सत्यमेव जयते

INDIAN AGRICULTURAL
RESEARCH INSTITUTE NEW DELHI

44418

I A R I 6
GIP NLK—H 3 I A R I —10 5 55—15,000

# PROCEEDINGS

OF THE

# NATIONAL ACADEMY OF SCIENCES

OF THE

UNITED STATES OF AMERICA

VOLUME 35, 1949

EXECUTIVE COMMITTEE
OF THE PROCEEDINGS

EDWIN BIDWELL WILSON, *Managing Editor*

R. A. MILLIKAN  D. W. BRONK

44418

*PUBLISHED MONTHLY*

Publication Office: Mack Printing Company, Easton, Pa.
Editorial Office: Harvard School of Public Health, Boston 15
Home Office of the Academy: Washington 25, D. C.

# CONTENTS

## NATIONAL ACADEMY OF SCIENCES

Page

National Academy of Sciences (Distribution and Ages of Membership) . . *By Fred. E. Wright, Home Secretary* 117

National Academy of Sciences—Organization . . . . . . . . . . . 423

National Academy of Sciences—The Conference on Bioelectric Potentials . . . . . . . . . . . . . . *By Duncan A. MacInnes* 547

## MATHEMATICS

A Theorem on Convex Bodies of the Brunn-Minkowski Type . . . . . *By Herbert Busemann* 27

The $n$-ality Theory of Rings . . . . . . . . . . *By Alfred L. Foster* 31

On a Theorem of Dieudonné . . . . . . . . . . *By Paul R. Halmos* 38

Osculating Conics of the Integral Curves of Third Order Differential Equations of the Type (*G*) . . *By Edward Kasner and John De Cicco* 43

Remark Concerning Integration . . . . . . . . . *By E. J. McShane* 46

Notes on Integration, IV . . . . . . . . . . . . . . *By M. H. Stone* 50

On the Degree of Convergence of Harmonic Polynomials to Harmonic Functions . . . . . *By J. L. Walsh, W. E. Sewell and H. M. Elliott* 59

A Simple Analytical Proof of a Fundamental Property of Birational Transformations . . . . . . . . . . . . . *By Oscar Zariski* 62

On Difference Sets . . . . . . . . . . . . . . *By S. Chowla* 92

Characters over Certain Types of Rings with Applications to the Theory of Equations in a Finite Field . . *By L. K. Hua and H. S. Vandiver* 94

Bayes Solutions of Sequential Decision Problems . . . . . . . . *By A. Wald and J. Wolfowitz* 99

Primary Ideals in Group Algebras . . . . . . . *By Irving Kaplansky* 133

Integral Representations of Bilinear Functionals. . . . . . . . *By Marston Morse and William Transue* 136

The Coefficients of Schlicht Functions, IV . . . . . . . . . . *By A. C. Schaeffer and D. C. Spencer* 143

On Zermelo's and von Neumann's Axioms for Set Theory . *By Hao Wang* 150

Bounds for the Solutions of a Second-Order Linear Differential Equation *By Walter Leighton* 190

Principal Quadratic Functionals and Self-Adjoint Second-Order Differential Equations . . . . . . . . . . . . *By Walter Leighton* 192

The Last Entry in Gauss's Diary . . . . . . . . . *By S. Chowla* 244

On the Theorem of Jentzsch . . . . . . . . . . *By Aryeh Dvoretzky* 246

On Deviations between Theoretical and Empirical Distributions . . . *By M. Kac* 252

A Theory of Trace-Admissible Algebras . . . . . . . *By A. A. Albert* 317

The Homotopy Groups of a Triad . *By A. L. Blakers and William S. Massey* 322

On the "Central" Probability Problem . . . . . . . *By Michel Loève* 328

On Congruences Which Relate the Fermat and Wilson Quotients to the Bernoulli Numbers . . . . . . . . . . . *By H. S. Vandiver* 332

Diffusion Equation and Stochastic Processes . . . . . *By S. Bochner* 368

On Epstein's Zeta Function (I) . . . . . . . *By S. Chowla and A. Selberg* 371

On a New Method in Elementary Number Theory Which Leads to an Elementary Proof of the Prime Number Theorem. . . . *By P. Erdös* 374

On the Structure of Locally Compact Groups . *By Andrew M. Gleason* 384

On the Automorphisms of a Sfield . . . . . . . . *By Loo-Keng Hua* 386

Lie Rings in Path Space . . . . . . . . . . . . *By D. D. Kosambi* 389

The Fréchet Variation and the Convergence of Multiple Fourier Series. *By Marston Morse and William Transue* 395

On the Zeros of Successive Derivatives of Entire Functions. . . . . *By Hans Rådström* 399

Homotopy Groups of Certain Algebraic Systems. . . . *By P. A. Smith* 405

Inequalities between the Two Kinds of Eigenvalues of a Linear Transformation. . . . . . . . . . . . . . . . *By Hermann Weyl* 408

Continued Fractions, Algebraic Functions and the Padé Table. . . . *By Richard Bellman and Ernst G. Straus* 472

On the Number of Solutions of Some Trinomial Equations in a Finite Field . . . . . . . . . *By Loo-Keng Hua and H. S. Vandiver* 477

On the Nature of the Solutions of Certain Equations in a Finite Field. *By L. K. Hua and H. S. Vandiver* 481

The Class Number of the Cyclotomic Field . . . . . . . . . . . *By N. C. Ankeny and S. Chowla* 529

Some Properties of a Sfield . . . . . . . . . . *By Loo-Keng Hua* 533

Imprimitivity for Representations of Locally Compact Groups I. . . . *By George W. Mackey* 537

Some Two-Person Games Involving Bluffing . . . . . . . . . . *By Richard Bellman and David Blackwell* 600

On Fluctuations in Coin-Tossing. . . *By Kai Lai Chung and W. Feller* 605

A Generalization of the Schwarz-Christoffel Transformation. . . . *By D. Gilbarg* 609

On Some Generalizations of the Cauchy-Frullani Integral . . . . . . *By A. M. Ostrowski* 612

On a Theorem of Weyl Concerning Eigenvalues of Linear Transformations. I . . . . . . . . . . . . . . . . *By Ky Fan* 652

On Self-Adjoint Differential Equations of Second Order . . . . . . *By Walter Leighton* 656

The Coefficient Problem for Bounded Schlicht Functions *By H. L. Royden* 657

Geometry of Reduced Moment Spaces . . . *By S. Karlin and L. S. Shapley* 673

On *n*-Dimensional Concepts of Bounded Variation, Absolute Continuity and Generalized Jacobian . . . *By T. Rado and P. V. Reichelderfer* 678

Quadratic Relations Involving the Numbers of Solutions of Certain Types of Equations in a Finite Field . . . . . *By H. S. Vandiver* 681

On the Use of Indeterminates in the Theory of Exponential Sums. . . *By H. S. Vandiver* 686

## ASTRONOMY

Some Problems of the Evolution of Stars. . . . . . *By Otto Struve* 161

## PHYSICS

Physical Systems of Curves in Space . . . . . . . . . . . . . *By Edward Kasner and John De Cicco* 106

Crystallographic Symmetry in Reciprocal Space. . . *By M. J. Buerger* 198

Physical Curves in Space of *n* Dimensions. . . . . . . . . . . . *By Edward Kasner and John De Cicco* 201

On Bending of Curved Thin-Walled Tubes . . . . *By Eric Reissner* 204
Note on the Theory of Lifting Surfaces . . . . . *By Eric Reissner* 208
Second Order Differential Equations of Rank 2 in Physics . . . . . . *By Edward Kasner and Don Mittleman* 338
Separation of the Variables in the One-Particle Schroedinger Equation in 3-Space . . . . . . . . . . *By Luther Pfahler Eisenhart* 412
Physical Families in Conservative Fields of Force . . . . . . . . . *By Edward Kasner and John De Cicco* 419
Superconductivity and the Bohr Magneton . . . . *By E. U. Condon* 488
Separation of the Variables of the Two-Particle Wave Equation . . . *By Luther Pfahler Eisenhart* 490

## ENGINEERING

First Approximation of Pressure Distribution on Curved Profiles at Supersonic Speeds . . . . . . . . . . . . *By T. Y. Thomas* 617

## CHEMISTRY

Aggregation Phenomena in Egg Albumin Solutions as Determined by Light Scattering Measurements . . . . . *By M. Bier and F. F. Nord* 17
Further Experiments with Vitamin A in Relation to Aging and to Length of Life . . . . . *By Henry C. Sherman and Helen Yarmolinsky Trupp* 90
The Preparation of Specific Adsorbents . . . . . *By Frank H. Dickey* 227
The Valence-State Energy of the Bivalent Oxygen Atom. . . . . . . *By Linus Pauling* 229
The Dissociation Energy of Carbon Monoxide and the Heat of Sublimation of Graphite . . . *By Linus Pauling and William F. Sheehan, Jr.* 359
Aggregation Phenomena in Polyvinyl Alcohol-Acetate Solutions. . . *By S. N. Timasheff, M. Bier and F. F. Nord* 364
On the Stability of the $S_8$ Molecule and the Structure of Fibrous Sulfur *By Linus Pauling* 495

## GEOLOGY

The Miocene Occurrence of Sequoia and Related Conifers in the John Day Basin . . . . . . . . . . . . . *By Ralph W. Chaney* 125
Early Tertiary Ecotones in Western North America *By Ralph W. Chaney* 356
The Principles of Time Series Analyses Applied to Ocean Wave Data. . *By H. R. Seiwell* 518

## BOTANY

Inhibition of Plant Growth by Protoanemonin and Coumarin, and Its Prevention by BAL . *By Kenneth V. Thimann and Walter D. Bonner, Jr.* 272
Antibiotic Substances from Basidiomycetes. IV. Marasmius Conigenus . *By Frederick Kavanagh, Annette Hervey and William J. Robbins* 343
A Method of Effecting Zygospore Germination in Certain Chlorophyceae *By Richard C. Starr* 453
The Correlation in Shape and Size Between Epidermal and Subepidermal Cells . . . . . . . . . . . . . . . . *By Frederic T. Lewis* 506

## ZOOLOGY

The Nucleoprotein Content of Whole Nuclei . . . . . . . . . . *By A. W. Pollister and C. Leuchtenberger* 66

Conjugation Between Double Monsters and Single Animals in Paramecium Bursaria . . . . . . . . . . . . *By Tze-Tuan Chen* 108

The Nature of the Specificity of Methyl Green for Chromatin. . . . *By Arthur W. Pollister and Cecilie Leuchtenberger* 111

The Culture of Paramecium Aurelia in the Absence of Other Living Organisms. . . . . *By W. J. van Wagtendonk and Patricia L. Hackett* 155

Responses of the Proximal Retinal Pigment of the Isolated Crustacean Eyestalk to Light and to Darkness . . . . . *By L. H. Kleinholz* 215

The Surface Gel Layer of Fundulus Eggs in Relation to Epiboly . . . *By J. P. Trinkaus* 218

On the Interpretation of Multi-Hit Survival Curves . . . . . . . . *By K. C. Atwood and A. Norman* 696

Atypical Pigment-Cell Differentiation in Embryonic Teleostean Grafts and Isolates . . . . . . . . . . . . *By Jane M. Oppenheimer* 709

## GENETICS

Unusual Gene-Controlled Combinations of Carbohydrate Fermentations in Yeast Hybrids . . . .*By Carl C. Lindegren and Gertrude Lindegren* 23

A Note on Centromere Organization . . . . . . . . . . . . . *By M. M. Rhoades and Warwick E. Kerr* 129

The Action of Allelic Forms of the Gene A in Maize II. The Relation of Crossing over to Mutation of $A^b$ . . . . *By John R. Laughnan* 167

Aberrant Heterozygotes in Escherichia Coli. . . .*By Joshua Lederberg* 178

The In Vitro Synthesis of Pantothenic Acid by Pantothenicless and Wild Type Neurospora . . . . . . . . . . . . *By Robert P. Wagner* 185

Acquisition of the J Substance by the Bovine Erythrocyte . . . . . . *By Clyde Stormont* 232

Further Studies on the Effect of Infra-Red Radiation on X-Ray-Induced Chromatid Aberrations in Tradescantia . . . . . *By C. P. Swanson* 237

Developmental Analysis of Genetic Differences in Pigmentation in the Axolotl . . . . . . . . . . . . . . . . . . *By H. Clark Dalton* 277

Inheritance of Carotenoid Differences in Lycopersicon Esculentum Strains . . . . . . . . . *By G. Mackinney and J. A. Jenkins* 284

The Beaded Minute Combination and Sex-Determination in Drosophila . *By A. H. Sturtevant* 311

The Intersexual Males of the Beaded Minute Combination in Drosophila Melanogaster . . . . . . . . . . . *By Richard B. Goldschmidt* 314

Chromosome Segregation in Maize Translocations in Relation to Crossing Over in Interstitial Segments . . . . . . *By C. R. Burnham* 349

Studies on the Biochemical Genetics of Yeast . . . . . . . . . . . *By Seymour Pomper and P. R. Burkholder* 456

Variation in the Amount of Desoxyribose Nucleic Acid in Different Tissues of Tradescantia . *By Franz Schrader and Cecilie Leuchtenberger* 464

Chromosomes of a Fox Hybrid (Alopex-Vulpes) . . . . . . . . . . *By Louise Wipf and Richard M. Shackelford* 468

THE EFFECT OF NITROGEN MUSTARD PRETREATMENT ON THE ULTRA-VIOLET-INDUCED MORPHOLOGICAL AND BIOCHEMICAL MUTATION RATE . . . . *By C. P. Swanson, W. D. McElroy and Helma Miller* 513

LINKAGE STUDIES OF THE RAT. X . . *By W. E. Castle and Helen Dean King* 545

THE RÔLE OF ORGANIC PEROXIDES IN THE INDUCTION OF MUTATIONS . . . . *By Frank H. Dickey, George H. Cleland and Carol Lotz* 581

CROSSING-OVER BETWEEN ALLELES AT THE LOZENGE LOCUS IN DROSOPHILA MELANOGASTER . . . . . . . . . *By M. M. Green and K. C. Green* 586

EXPERIMENTS ON LIGHT-REACTIVATION OF ULTRA-VIOLET INACTIVATED BACTERIA . . . . . . . . . . . . . *By A. Novick and Leo Szilard* 591

A GENIC SERIES CONTROLLING CHLOROPLAST PIGMENTS' PRODUCTION IN DIPLOID ZEA MAYS L. . . . . . . . . . . . *By Herbert L. Everett* 628

REPRODUCTION AND DISTRIBUTION OF THE CYTOPLASMIC FACTOR FOR MALE STERILITY IN MAIZE . . . . . . . . . . *By Warren H. Gabelman* 634

THE EFFECT OF OXYGEN ON THE FREQUENCY OF X-RAY INDUCED CHROMOSOMAL REARRANGEMENTS IN TRADESCANTIA MICROSPORES . . . . . . . . *By Norman H. Giles, Jr., and Herbert Parkes Riley* 640

THE INFLUENCE OF THE Y-CHROMOSOME ON X-RAY INDUCED MUTATION RATE IN DROSOPHILA MELANOGASTER . . . . . . . *By Jean Kerechner* 647

## PHYSIOLOGY AND BIOCHEMISTRY

RIBOFLAVIN-SENSITIZED PHOTOÖXIDATION OF INDOLEACETIC ACID AND RELATED COMPOUNDS . . . . . . . . . . . . . *By Arthur W. Galston* 10

THE MODE OF ACTION OF THE METAL-PEPTIDASES . . . *By Emil L. Smith* 80

INDIVIDUAL METABOLIC PATTERNS, ALCOHOLISM, GENETOTROPHIC DISEASES . . *By Roger J. Williams, L. Joe Berry and Ernest Beerstecher, Jr.* 265

EVIDENCE FOR A TRYPTOPHANE CYCLE IN NEUROSPORA . . . . . . . . *By Francis A. Haskins and Herschel K. Mitchell* 500

SOME BIOELECTRICAL PROBLEMS . . . . . . . . *By W. J. V. Osterhout* 548

SOME PHYSICAL ASPECTS OF BIOELECTRIC PHENOMENA . . *By Kenneth S. Cole* 558

THE SOURCE OF THE BIOELECTRIC POTENTIALS IN LARGE PLANT CELLS . . *By L. R. Blinks* 566

QUINOLINIC ACID ACCUMULATION IN THE CONVERSION OF 3-HYDROXYANTHRANILIC ACID TO NIACIN IN NEUROSPORA *By David M. Bonner and Charles Yanofsky* 576

## PATHOLOGY AND BACTERIOLOGY

THE ISOLATION OF BIOCHEMICALLY DEFICIENT MUTANTS OF BACTERIA BY MEANS OF PENICILLIN . . . . . . . . . . . . . *By Bernard D. Davis* 1

EFFECT OF VISIBLE LIGHT ON THE RECOVERY OF STREPTOMYCES GRISEUS CONIDIA FROM ULTRA-VIOLET IRRADIATION INJURY . . . . . . *By Albert Kelner* 73

PHARMACOLOGICAL MODIFICATION OF RESISTANCE TO RADIATION—A PRELIMINARY REPORT . . . . . *By John B. Graham and Ruth M. Graham* 102

THE MECHANISM OF PROTEINURIA . . . . . . . . . . . *By T. Addis* 194

AEROBIC FORMATION OF FUMARIC ACID IN THE MOLD RHIZOPUS NIGRICANS: SYNTHESIS BY DIRECT $C_2$ CONDENSATION . . *By J. W. Foster, S. F. Carson, D. S. Anthony, J. B. Davis, W. E. Jefferson and M. V. Long* 663

## ANTHROPOLOGY AND PSYCHOLOGY

ON INTERPOSITION AS A CUE FOR THE PERCEPTION OF DISTANCE *By P. Ratoosh* 257

VALUES AS A SUBJECT OF NATURAL SCIENCE INQUIRY . . . *By A. L. Kroeber* 261

INFRARED ABSORPTION IN FIELD STUDIES OF OLFACTION IN HONEYBEES . . .
*By Walter R. Miles and Lloyd H. Beck* 292
SOME RECENT APPROACHES TO THE EXPERIMENTAL STUDY OF HUMAN FATIGUE.
*By Leonard Carmichael, John L. Kennedy and Leonard C. Mead* 691

## ERRATA

SOME PROPERTIES OF ROTATIONAL FLOW OF A PERFECT GAS . . . . . . .
*By P. Neményi and R. Prim* 116
BOUNDS FOR THE SOLUTIONS OF A SECOND-ORDER LINEAR DIFFERENTIAL EQUATION
*By Walter Leighton* 422
AGGREGATION PHENOMENA IN POLYVINYL ALCOHOL-ACETATE SOLUTIONS . . .
*By S. N. Timasheff, M. Bier and F. F. Nord* 494
STUDIES ON THE BIOCHEMICAL GENETICS OF YEAST . . . . . . . . . . .
*By Seymour Pomper and P. R. Burkholder* 712

The Effect of Nitrogen Mustard Pretreatment on the Ultra-Violet-Induced Morphological and Biochemical Mutation Rate . . . .
*By C. P. Swanson, W. D. McElroy and Helma Miller* 513
Linkage Studies of the Rat. X . . *By W. E. Castle and Helen Dean King* 545
The Rôle of Organic Peroxides in the Induction of Mutations . . . .
*By Frank H. Dickey, George H. Cleland and Carol Lotz* 581
Crossing-Over Between Alleles at the Lozenge Locus in Drosophila Melanogaster . . . . . . . . *By M. M. Green and K. C. Green* 586
Experiments on Light-Reactivation of Ultra-Violet Inactivated Bacteria . . . . . . . . . . . . . *By A. Novick and Leo Szilard* 591
A Genic Series Controlling Chloroplast Pigments' Production in Diploid Zea Mays L. . . . . . . . . . . . *By Herbert L. Everett* 628
Reproduction and Distribution of the Cytoplasmic Factor for Male Sterility in Maize . . . . . . . . . . *By Warren H. Gabelman* 634
The Effect of Oxygen on the Frequency of X-Ray Induced Chromosomal Rearrangements in Tradescantia Microspores . . . . . . . . .
*By Norman H. Giles, Jr., and Herbert Parkes Riley* 640
The Influence of the Y-Chromosome on X-Ray Induced Mutation Rate in Drosophila Melanogaster . . . . . . . *By Jean Kerechner* 647

## PHYSIOLOGY AND BIOCHEMISTRY

Riboflavin-Sensitized Photoöxidation of Indoleacetic Acid and Related Compounds . . . . . . . . . . . . . *By Arthur W. Galston* 10
The Mode of Action of the Metal-Peptidases . . . *By Emil L. Smith* 80
Individual Metabolic Patterns, Alcoholism, Genetotrophic Diseases . .
*By Roger J. Williams, L. Joe Berry and Ernest Beerstecher, Jr.* 265
Evidence for a Tryptophane Cycle in Neurospora . . . . . . . . .
*By Francis A. Haskins and Herschel K. Mitchell* 500
Some Bioelectrical Problems . . . . . . . . *By W. J. V. Osterhout* 548
Some Physical Aspects of Bioelectric Phenomena . . *By Kenneth S. Cole* 558
The Source of the Bioelectric Potentials in Large Plant Cells . .
*By L. R. Blinks* 566
Quinolinic Acid Accumulation in the Conversion of 3-Hydroxyanthranilic Acid to Niacin in Neurospora *By David M. Bonner and Charles Yanofsky* 576

## PATHOLOGY AND BACTERIOLOGY

The Isolation of Biochemically Deficient Mutants of Bacteria by Means of Penicillin . . . . . . . . . . . . . . *By Bernard D. Davis* 1
Effect of Visible Light on the Recovery of Streptomyces Griseus Conidia from Ultra-violet Irradiation Injury . . . . . . *By Albert Kelner* 73
Pharmacological Modification of Resistance to Radiation—A Preliminary Report . . . . . . *By John B. Graham and Ruth M. Graham* 102
The Mechanism of Proteinuria . . . . . . . . . . . *By T. Addis* 194
Aerobic Formation of Fumaric Acid in the Mold Rhizopus Nigricans: Synthesis by Direct $C_2$ Condensation . . . *By J. W. Foster, S. F. Carson, D. S. Anthony, J. B. Davis, W. E. Jefferson and M. V. Long* 663

## ANTHROPOLOGY AND PSYCHOLOGY

On Interposition as a Cue for the Perception of Distance *By P. Ratoosh* 257
Values as a Subject of Natural Science Inquiry . . . *By A. L. Kroeber* 261

INFRARED ABSORPTION IN FIELD STUDIES OF OLFACTION IN HONEYBEES . . .
*By Walter R. Miles and Lloyd H. Beck* 292
SOME RECENT APPROACHES TO THE EXPERIMENTAL STUDY OF HUMAN FATIGUE.
*By Leonard Carmichael, John L. Kennedy and Leonard C. Mead* 691

## ERRATA

SOME PROPERTIES OF ROTATIONAL FLOW OF A PERFECT GAS . . . . . . .
*By P. Neményi and R. Prim* 116
BOUNDS FOR THE SOLUTIONS OF A SECOND-ORDER LINEAR DIFFERENTIAL EQUATION
*By Walter Leighton* 422
AGGREGATION PHENOMENA IN POLYVINYL ALCOHOL-ACETATE SOLUTIONS . . . .
*By S. N. Timasheff, M. Bier and F. F. Nord* 494
STUDIES ON THE BIOCHEMICAL GENETICS OF YEAST . . . . . . . . . . .
*By Seymour Pomper and P. R. Burkholder* 712

PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES
Volume 35 January 15, 1949 Number 1

# *THE ISOLATION OF BIOCHEMICALLY DEFICIENT MUTANTS OF BACTERIA BY MEANS OF PENICILLIN*

BY BERNARD D. DAVIS*

FROM THE U. S. PUBLIC HEALTH SERVICE, TUBERCULOSIS RESEARCH LABORATORY, CORNELL UNIVERSITY, MEDICAL COLLEGE, NEW YORK

Communicated by R. J. Dubos, November 18, 1948

It is a simple matter to isolate bacterial mutants when the mutants can proliferate or survive in an environment which suppresses or eliminates the parent strain. There is consequently no difficulty in obtaining mutants, even of low frequency, which differ from the parent strain by being resistant to antibacterial chemicals or viruses, or by having decreased nutritional requirements. Mutants with increased nutritional requirements, however, have been much less convenient to isolate. Recently developed techniques [1, 2] permit a considerable improvement over the earlier practice of random selection, but still permit selection from only a few hundred colonies per agar plate. This paper is concerned with a method of obtaining biochemically deficient mutants from very much larger populations. The method is based on the unusual mode of action of penicillin, which sterilizes only growing bacteria. Mutants which are unable to grow in minimal medium therefore survive, while the predominant non-mutant population is sterilized.

*Historical Introduction.*—The systematic isolation of microbial mutants with increased nutritional requirements was initiated by Beadle and co-workers,[3, 4, 5] using the mold *Neurospora crassa*, and has since been extended by others to *Escherichia coli* and other bacteria (cf. 6), and to other molds. With respect to the genetic interpretation of the mutations, Neurospora has the advantage of permitting either sexual fusion or asexual multiplication at will; with bacteria, on the other hand, recombination of genetic characters, which has been discovered only recently,[7, 8] is apparently a rare event and is demonstrable with few bacterial strains. From the point of view of their use as biochemical tools, however, mutants of bacteria have certain advantages, especially with respect to ease of handling. Their rapid multiplication leads to maximal growth in 24 hours or less, and their growth as discrete cells and isolated colonies, in contrast to the mycelial mat of Neurospora, permits precise quantitative treatment of populations of any size. Bacterial numbers below the visible level can be measured by plating out on solid media, while denser growth can be measured most simply by turbidimetry. The possibility of aerobic or anaerobic growth increases the variety of problems which can be attacked. In addition, the uniform production of colonies and streaks on solid media permits easy recognition of the production by one bacterial strain of a growth factor which diffuses through the agar to stimulate another strain streaked nearby.[9] Finally, as will be described below, it has become possible to isolate bacterial mutants considerably more efficiently than has been possible with molds, although recent developments may also render the isolation of mold mutants quite efficient.[10, 11]

The analysis of the behavior of deficient mutants, so far studied mostly with Neurospora, represents one of the major developments in biochemistry of the past decade. It has permitted recognition of a number of distinct enzymic steps in the course of various biosyntheses—a method of exceptional interest as the attention of biochemistry is increasingly focused on anabolic reactions. This development may also be expected to have several important applications, including microbiological assay; the recognition of previously unknown metabolites which can serve as models for synthesis of chemotherapeutic analogs; and perhaps microbial production of rare biochemicals, since certain deficient mutants have been shown to accumulate the substrate of the absent enzyme.[12]

In particular, illustrating the unity of cellular metabolism,[13, 14, 15] the growth factors (nutrilites) of various microorganisms found in nature have almost without exception turned out to be essential participants in the intermediary metabolism of higher animals and plants. In a number of instances the microorganisms have provided the first means of recognizing and isolating vitamins (essential growth factors of animals). But since the growth requirements of a large proportion of "wild type" organisms have already been identified, it appears reasonable to suppose that in the future the discovery of unknown metabolites will increasingly depend on the production of mutants for which these metabolites are nutrilites.

Under these circumstances, it would be highly desirable to have a technique for isolating biochemical mutants of microorganisms more efficiently than has hitherto been possible. The work on Neurospora has been done with mutants obtained by testing large numbers of unselected individual spores. The total frequency of detectable biochemically deficient mutants following irradiation is reported to be about 2% with Neurospora;[4] maximal values of 6% have been recorded with bacteria.[16] The incidence of any given type of mutant is extremely small. With this method it would not be profitable to undertake to isolate specifically any single type of mutant which might be wanted by a biochemist, nor could one expect to obtain, except by rare chance, those mutants whose frequency of appearance following irradiation is less than perhaps $10^{-4}$, and whose existence can at present only be postulated.

A considerable improvement in efficiency of isolating biochemically deficient bacterial mutants has been effected by Lederberg,[1] based on the delayed production of colonies by the mutants as a result of delaying the enrichment of the minimal agar medium until the non-mutants have grown for a while. A related method has recently been described which involves the production of small colonies of mutants by limiting the amount of enrichment.[2] Both these methods, however, still restrict the search to comparatively common mutants, since a maximum of approximately 500 bacterial colonies per plate can be studied.

*Principle of the Penicillin Method.*—The possibility, with biochemically deficient mutants, of selectively eliminating a large non-mutant population suggested itself on the basis of the reports[17, 18] that penicillin sterilizes only multiplying bacteria. The data on which these reports were based show only that penicillin failed to sterilize bacteria in a grossly deficient menstruum, such as physiological saline, or in an adequate medium kept in the refrigerator. In order to test whether penicillin would also fail to sterilize a deficient mutant in a medium which was adequate except for the specific deficiency, the present investigation was begun by isolating a tryptophane-requiring mutant of *E. coli* by one of the earlier methods.[2] This mutant was not sterilized by penicillin in a minimal medium lacking tryptophane, while the parent stock strain was rapidly sterilized by penicillin in this medium (and the mutant in a medium containing tryptophane). It therefore appeared that penicillin should be an ideal selective agent for sterilizing large non-mutant populations in a minimal medium, while permitting survival of those mutants which fail to grow on this medium. Accordingly, a search was made for biochemical mutants among the survivors of extensive but incomplete sterilization by penicillin.

Experiments carried out over a period of nearly a year with several bacterial species consistently failed to yield any mutants among the survivors of irradiated populations exposed to penicillin in minimal medium. Success was attained, however, when the bacteria were passed through an extra stage of cultivation in enriched medium after irradiation, followed by washing to eliminate the growth factors present in the enriched medium. They were then suspended in minimal medium containing penicillin and finally plated in enriched agar media. This modified procedure yielded a large number of survivors (up to a hundred per plate, from $10^5$ organisms exposed to penicillin). These colonies consisted practically exclusively (90 to 100%) of a variety of strains unable to grow on minimal medium.

The introduction of the extra stage of cultivation was based on the hypothesis that the previous failures were caused by the syntrophic effect of metabolites released by the relatively dense populations of bacteria exposed to penicillin following irradiation. This hypothesis was supported by experiments on mixtures of the washed tryptophane-less mutant and the stock culture; these showed that the mutant became increasingly susceptible to penicillin in minimal medium when the density of stock bacteria, either normal or sterilized by irradiation, exceeded $10^6$ cells per milliliter. Since in the early experiments the suspensions used as sources of mutants were irradiated until the proportion of viable survivors was $10^{-4}$ to $10^{-5}$, it was necessary, in order to expose appreciable numbers of viable bacteria to penicillin, that the total population density be very high. The extra stage of cultivation eliminated the harmful excess of metabolically active but non-viable products of irradiation.

Following this adaptation of the method, it was found that the basis on which the extra cultivation had been introduced was not correct after all, even though the procedure was successful. Once the irradiated bacteria had passed through the intermediate cultivation, a certain number of mutants could be obtained even at very high population densities. It was therefore impossible to account for the requirement of intermediate cultivation simply on the basis of the syntrophic effect. We believe the main factor is rather the following. Mutations of the desired type are produced by eliminating the function of a gene which normally produces molecules of a given species of enzyme; these enzyme molecules in turn control a step in the biosynthesis of a given metabolite. But when the gene has undergone a mutation following irradiation, even though the product of the gene can no longer be formed, the pre-existing gene products (enzyme molecules) continue to function and to synthesize their metabolites. The cell is therefore able to grow to a certain extent on minimal medium, and is consequently sterilized by penicillin before the products of the mutated gene are exhausted. During the process of intermediate cultivation, however, the original gene products are so diluted out or exhausted that the mutants are unable to grow on minimal medium, and hence are insensitive to penicillin. To describe this "phenotypic lag" more briefly one might say that the new genotype finds phenotypic expression as a biochemically deficient mutant only after it has undergone sufficient growth (possibly requiring several generations) to lose the physiologic character of the old genotype. The syntrophic effect appears also to play a definite but secondary rôle in determining survival from penicillin. Experiments are in progress to analyze the syntrophic effect and the phenotypic lag quantitatively.

By this technique a variety of mutants of *E. coli* ("Waksman" strain, ATCC #9637) have been obtained with individual requirements for all the naturally occurring amino acids except alanine, aspartic acid and hydroxyproline; for several multiple sets of amino acids; for the naturally occurring purines or their nucleosides or nucleotides; for the naturally occurring pyrimidines or their nucleosides or nucleotides; for the vitamins biotin, niacin, *p*-amino benzoic acid, thiamin, pyridoxin and panthothenic acid; and for unidentified factors in yeast extract. Mutants requiring vitamins have been isolated with less efficiency than the others; we have evidence that the syntrophic effect is much more striking with at least some of the vitamins, which act as nutrilites in much smaller traces, than with amino acids or nucleic acid components, which form the bulk of the protoplasm.

The experimental details underlying this paper will be published later. A brief protocol of a typical experiment follows, in which only amino acid mutants were sought.

*Experimental.*—A 24-hour turbid culture of *E. coli* ("Waksman" strain, ATCC #9637) in minimal medium [19] was irradiated for 2 minutes in a quartz flask with constant shaking at a distance of 39 inches from an ultraviolet mercury lamp (General Electric "Sterilamp"). This procedure reduced the viable count from $10^8$ per milliliter to $10^7$ per milliliter. (In most experiments the irradiation was more extensive.) One milliliter of the suspension was added to 2 ml. of minimal medium supplemented with 0.2% tryptic hydrolyzate of casein (Sheffield "N-Z-Case"), and the culture was incubated for 24 hours at 37°C., at which time the viable count had reached $1.5 \times 10^9$ per milliliter. The bacteria were centrifuged, washed once with water and resuspended in 3 ml. of water. Serial tenfold dilutions of the suspension were prepared in water, and 0.1 ml. of each dilution added to 3 ml. of minimal medium containing 300 units of crystalline penicillin per milliliter. Following incubation at 37°C. for 24 and 48 hours, 0.1 ml. from each tube was plated in 10 ml. of minimal agar and in the same medium supplemented with 0.2% casein hydrolyzate. (It was determined that this dilution of the inoculum in agar was sufficient to remove the antibacterial effect of the penicillin.) Because of the frequently delayed appearance of colonies following exposure to penicillin, the plates were incubated for 48 rather than 24 hours. Table 1 shows the much larger numbers of colonies appearing in the enriched medium. Ten colonies chosen at random from one of the plates of enriched medium were all found to be biochemically deficient mutants, unable to grow on minimal agar. The mutant strains were spot-tested [22] for their response to all of the naturally occurring amino acids. Because of the multiplication during the intermediate cultivation between irradiation and selection by penicillin, a number of replicate colonies with the same requirement, presumably a clone derived from a single mutation, are generally found in the same plate. In a typical experiment 5 to 10 distinct types are found on a single plate.

*Discussion.*—*E. coli* was the organism chosen because of its hardiness, its ability to grow on the simplest medium, its completely dispersed growth in liquid medium, and the large amount of information already available on its biochemical mutants. Though it requires much higher concentrations of penicillin than those species which are considered fully sensitive to the drug, it fortunately shares with these species the requirement of bacterial growth for bactericidal action. With more sensitive species the separation of mutants from non-mutants by penicillin might well be even more quantitative.

It may be emphasized that those mutants which have survived penicillin under these special conditions are not penicillin resistant in the ordinary sense. In the presence of their growth requirements they display no resistance to the bactericidal action of penicillin. It would not be expected that biochemical deficiencies should be associated with resistance to penicillin; on the contrary, it has been pointed out that the development of a high degree of penicillin resistance by *Staphylococcus aureus* is accompanied by loss of most of the growth requirements of this normally fastidious organism.[28]

TABLE 1

SURVIVAL OF MUTANTS AND NON-MUTANTS AFTER EXPOSURE TO PENICILLIN (300 U./ML.) IN MINIMAL MEDIUM

| | COLONIES IN PLATING OF 0.1 ML. | | | |
|---|---|---|---|---|
| | *AFTER 24 HRS. IN PENICILLIN* | | *AFTER 48 HRS. IN PENICILLIN* | |
| INOCULUM IN 3-ML. TUBE | MINIMAL MEDIUM | ENRICHED MEDIUM | MINIMAL MEDIUM | ENRICHED MEDIUM |
| $10^8$ bacteria | ca. 400 | ca. 400 | 12 | 51 |
| $10^7$ | 36 | 152 | 1 | 67 |
| $10^6$ | 7 | 79 (10/10 mut.) | 0 | 36 |

So far as is known to the author, penicillin is the only antibacterial agent which has been shown to require growth for its bactericidal action, but few others have been tested as carefully. Streptomycin[9] and sulfonamides have other modes of action. For microbial species which are completely resistant to penicillin it might be worth while to search for at least partial selection of mutants by other antimicrobial agents. It is well known that a variety of disinfectants sterilize cells from the logarithmic phase of growth more rapidly than cells from the stationary phase.

It was incidentally anticipated, in undertaking this investigation, that light might be thrown on the mode of action of penicillin, since a selective survival of certain classes of mutants, but not of others, would have furnished an indication of the site of action of the drug. As it turns out, the apparently universal survival of all types of non-multiplying mutants makes this procedure particularly valuable as a tool for selecting mutants, but valueless for narrowing down the search for the point of metabolic attack of penicillin. One can conclude only that sterilization by this compound requires extensive protoplasmic synthesis—an integrated process which apparently is interrupted at all points (including the point or points of attack of penicillin) soon after a deficiency appears at any one point. It may be pointed out, however, that actual cell division is not required. Considerable sterilization by penicillin occurs during the lag phase, before the number of cells has increased in a parallel culture without penicillin.[9]

This method as developed so far does not appear to yield quantitative survival of mutants, and hence cannot be used for accurate determination of mutation frequencies. The total fraction of mutants obtained with casein hydrolyzate as a supplement is approximately $10^{-3}$ of the viable cells exposed to penicillin, whereas a frequency of approximately $10^{-2}$ was obtained when the same irradiated suspension was tested by another method[2] which permits survival of all the members of a small population. The virtue of the penicillin method lies chiefly in the use it permits of huge populations. It is possible by this method to isolate certain specific types of mutants conveniently. For example, when it turned out that all of the cystine-requiring mutants on hand responded to sulfide, an experiment was performed to isolate mutants blocked at another level. Cystine was added in the intermediate cultivation, and sulfide in the penicillin tube to eliminate the sulfide mutants. Subsequent plating yielded several kinds of cystine mutants which did not respond to sulfide.

The relatively high frequency of several per cent of detectable biochemical deficiencies in irradiated bacteria has been interpreted by Burkholder and Giles[16] as implying a haploid, uninucleate state of the bacterial cell. Otherwise, it was suggested, the heterozygous sets of genes in the same cell would lead to mixed colonies, in which the unmutated strain would regularly have prevented detection of mutants by the methods in use, which involved formation of a colony from each cell immediately following irradiation. Although this argument was advanced in connection with *Bacillus subtilis* spores, it would apply equally, if valid, to other bacteria, which have similar mutation frequencies following ultra-violet irradiation. Actually, this interpretation is in conflict with recent cytological evidence that bacilli, including *E. coli*, are multinucleate.[24] Although the residence in these nuclear bodies of the genes for biosyntheses has not been proved, it seems very probable. The argument of Burkholder and Giles, in any case, is open to question if we recall that the irradiated suspensions do not contain 1% mutants and 99% non-mutants; rather, in a typical case, with $10^{-4}$ viable survivors, the proportions would be $10^{-6}$ mutants, $99 \times 10^{-6}$ non-mutants, and 0.9999 non-viable cells which may be regarded in a general way as lethal mutants. It is obvious that under these circumstances a viable mutant set of genes in a multinucleate cell would almost invariably be accompanied by a lethal set which could not give rise to a mixed colony, and presumably would not interfere with the propagation of the viable set of genes. This consideration is also applicable to the suggestion [25, 26] that polyploidy might be a cause of the delayed appearance of phage-resistant bacterial mutants.

Delayed phenotypic expression appears to be much the most likely, though not the only possible, explanation of the failure to isolate mutants by exposure to penicillin immediately after irradiation; for it seems not only plausible, but practically inevitable, that enzymes should persist in the cell for some time after radiation has altered the corresponding gene. It is not permissible to refer to this process as a "cytoplasmic lag," such as has been demonstrated with Paramecium,[27] for we have no certain knowledge of the location in bacteria of either the genes or their products. It would be desirable, however, for the biochemical analysis of genetic processes, to have a collective term for the products of gene action, defined on a functional rather than a morphological basis. *Genome* has long been used to denote the sum total of the autoreproductive units (genes) in a cell. We here propose the term *phenome*[28] for the sum total of extragenic, non-autoreproductive portions of the cell, whether cytoplasmic or nuclear. The phenome would be the material basis of the phenotype, just as the genome is the material basis of the genotype. What has been called a delay in phenotypic expression would, then, depend on the transformation of the phenome to correspond to the new genome. Since the word phenotype is purely formal and has somewhat special implications as ordinarily used in genetics, it is suggested that the process under discussion be referred to more concretely as a *phenomic lag*.

A delay in the appearance of mutants induced by irradiation has also been described by Demerec for phage resistance.[25, 26] Newcombe[29] has recently demonstrated indirectly a similar delay among spontaneously appearing phage-resistant mutants, and favors the view that in both induced and spontaneous phage resistance the delay is one of phenotypic expression of the mutation. Biochemically deficient mutants are a simpler case in the sense that we have some conception of the biochemical nature of the change, whereas the mechanism underlying phage resistance is more obscure. The similar delay in phenotypic expression of both types of mutation therefore supports the notion that in phage resistance there is also an alteration of the composition of the phenome, presumably primarily involving one or more enzymes controlled by a single gene.

*Summary*.—A method is described for the isolation of biochemically deficient mutants of bacteria. This procedure is based on the fact that penicillin sterilizes only growing bacteria; mutants selectively survive exposure to penicillin in a minimal medium which is inadequate for their nutrition. By this technique a large variety of mutants of *E. coli* have been obtained with requirements for amino acids, nucleic acid components, and vitamins.

Mutants induced by ultra-violet irradiation do not survive exposure

to penicillin in minimal medium unless they are first permitted a period of growth, during which the new genotype achieves physiologic expression as a deficient phenotype. The term *phenome* is proposed to denote the total extragenic material of the cell, and the term *phenomic lag* to describe the delay in phenotypic expression.

Syntrophism also contributes to the susceptibility of mutants to penicillin in minimal medium.

The excellent technical assistance of Mrs. Harlean Cort is gratefully acknowledged.

---

[*] The use of penicillin to isolate biochemically deficient mutants was also developed independently by J. Lederberg and N. Zinder. Communications by these investigators and by the author are being published in the December 1948 issue of the *Journal of the American Chemical Society.*

[1] Lederberg, J., and Tatum, E. L., *J. Biol. Chem.*, **165**, 381 (1946).

[2] Davis, B. D., *Arch. Biochem.*, in press.

[3] Beadle, G. W., and Tatum, E. L., PROC. NAT. ACAD. SCI., **27**, 499 (1941).

[4] Beadle, G. W., *Physiol. Rev.*, **25**, 643 (1945).

[5] Beadle, G. W., *Ann. Rev. Physiol.*, **10**, 17 (1948).

[6] Tatum, E. L., *Cold Spring Harbor Symp. Quant. Biol.*, **11**, 278 (1946).

[7] Tatum, E. L., and Lederberg, J., *J. Bact.*, **53**, 673 (1947).

[8] Lederberg, J., *Genetics*, **32**, 505 (1947).

[9] Davis, B. D., unpublished observations.

[10] Fries, N., *Nature*, **159**, 199 (1947).

[11] Lein, J., Mitchell, H. K., and Houlahan, M. B., PROC. NAT. ACAD. SCI., **34**, 435 (1948).

[12] Horowitz, N. H., *J. Biol. Chem.*, **162**, 413 (1946).

[13] Lwoff, A., *L'Evolution Physiologique*, Hermann & Cie., Paris, 1943.

[14] Mueller, J. H., *Harvey Lect.*, **39**, 143 (1943–1944).

[15] Knight, B. C. J. G., *Vit. & Hor.*, **3**, 105 (1945).

[16] Burkholder, P. R., and Giles, N. H., Jr., *Am. J. Bot.*, **34**, 345 (1947).

[17] Hobby, G. L., Meyer, K., and Chaffee, E., *Proc. Soc. Exp. Biol. & Med.*, **50**, 281 (1942).

[18] Chain, E., and Duthie, E. S., *Lancet*, **1**, 652 (1945).

[19] The minimal medium had the following composition: $Na_2HPO_4 \cdot 7H_2O$ 6 g.; KCl 4 g.; $K_2SO_4$ 1 g.; $NH_4$ lactate 2 g.; $MgSO_4 \cdot 7H_2O$ 0.025 g.; glucose (autoclaved separately) 2 g.; $H_2O$ ad 1000 ml.; pH adjusted to 7.6. This medium contains considerably higher concentrations of K and Mg than certain other media[20] which have been recommended for *E. coli*, including Medium "F" which is widely used in bacteriophage work.[21] It was found that these high concentrations are necessary for the initiation of growth by small inocula (10 to $10^4$ cells) and for the efficient sterilization of non-mutants by penicillin, although the extra K and Mg do not affect the final density of growth attained with large inocula. A small amount of sodium citrate (0.01%) was also found to substitute for the high concentrations of K and Mg, and was sometimes added to the minimal medium since it augmented the bactericidal rate of penicillin.

[20] MacLeod, C. M., *J. Exp. Med.*, **72**, 217 (1940).

[21] Cohen, S. S., and Anderson, T. F., *Ibid.*, **84**, 511 (1946).

[22] Lederberg, J., *J. Bact.*, **52**, 503 (1946).

[23] Gale, E. F., and Rodwell, A. W., *Ibid.*, **55**, 161 (1948).

[24] Robinow, C. F., addendum to Dubos, R. J., *The Bacterial Cell*, Harvard Univ. Press, Cambridge, Mass., 1945.

[25] Demerec, M., PROC. NAT. ACAD. SCI., **32,** 36 (1946).

[26] Demerec, M., and Latarjet, R., *Cold Spring Harbor Symp. Quant. Biol.,* **11,** 38 (1946).

[27] Sonneborn, T. M., *Genetics,* **28,** 90 (1943).

[28] The author is indebted to Dr. Werner Maas for assistance in selecting the neologism "phenome."

[29] Newcombe, H. B., *Genetics,* **33,** 447 (1948).

---

# *RIBOFLAVIN-SENSITIZED PHOTOÖXIDATION OF INDOLEACETIC ACID AND RELATED COMPOUNDS*

BY ARTHUR W. GALSTON*

KERCKHOFF LABORATORIES OF BIOLOGY, CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated by G. W. Beadle, November 18, 1948

Indoleacetic acid (IAA) is known to exert profound effects upon growth and morphogenesis in higher plants (cf., the review by Went and Thimann[17]). Since its natural occurrence in plants has been directly demonstrated by chemical isolation (Haagen-Smit, *et al.,*[7, 8]) its status as a plant hormone seems well established. Obviously, any information concerning the genesis and disappearance of this hormone is of fundamental importance to plant physiology.

During a series of experiments on the physiology of light action in etiolated peas (Galston and Hand[5]), we noticed that the addition of small amounts of riboflavin (Rbf) to the growth medium resulted in a marked growth inhibition if the tissue were exposed to light. No such inhibition occurred in the absence of light. Since this inhibition could be partially reversed by the addition of relatively large quantities of IAA to the medium, it appeared possible that riboflavin in some way caused the photo-inactivation of indoleacetic acid. This interpretation was completely confirmed by *in vitro* experiments, as described below.

*Methods.*—The test solutions were put into Erlenmeyer flasks of such capacity that a layer 1–2 cm. deep was formed. Duplicate series were prepared, one being mixed and stored in a dark room, the other being exposed to about 200 foot-candles of light from "Daylight" fluorescent bulbs. Such light as was necessary in the "darkroom" was supplied by a $7^1/_2$-watt ruby-red bulb. In the determination of IAA, a 1-cc. aliquot of the reaction mixture was removed to a test tube, mixed with 4 cc. of Salkowski reagent (see Tang and Bonner[14]), and allowed to stand for 30 minutes. The intensity of the resultant pink color was then measured in a Klett-Summerson photoelectric colorimeter, using a green filter, and the concentration of IAA in the aliquot determined by reference to a previously prepared standard curve.

The riboflavin employed was the Merck product, made up into a stock solution of 100 γ/cc. and stored in a red bottle in the refrigerator to retard decomposition. The IAA utilized was made up into a stock solution of 50 γ/cc., adjusted to pH 7.0 and similarly stored in the refrigerator until use.

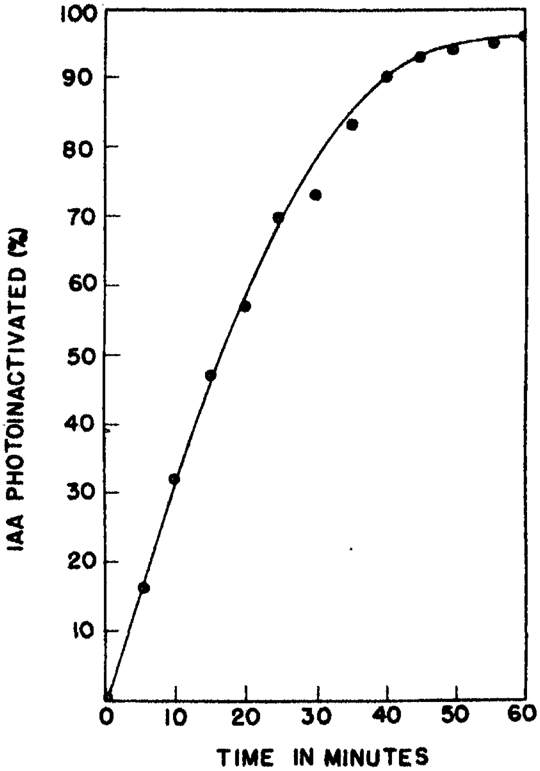


FIGURE 1

The time course of photoinactivation of indoleacetic acid. Solution contained 25 γ/cc. IAA and 10 γ/cc. Rbf.

*Experiments.*—Although solutions of indoleacetic acid are quite stable both in light and in the dark, the addition of as little as 0.01 γ/cc. of riboflavin renders the IAA susceptible to rapid light inactivation. Thus, in a solution containing 25 γ/cc. IAA and 1 γ/cc. riboflavin, practically all of the indoleacetic acid is inactivated within 1 hour. Typical data are presented in table 1.

To determine the time course and rate of the reaction, 25 γ/cc. IAA and 10 γ/cc. of riboflavin were incubated at 23°C. under the fluorescent lights.

Every five minutes, a 1-cc. aliquot was removed from the solution and subjected to colorimetric analysis. The data (Fig. 1) show that the reaction is 50% complete in about 16 minutes, 75% complete in about 31 minutes, and virtually complete at the end of one hour.

TABLE 1

THE EFFECT OF RIBOFLAVIN AND OF LIGHT ON THE DISAPPEARANCE FROM SOLUTION OF IAA. DURATION OF EXPERIMENT, 1 HOUR; TEMPERATURE, 23°C.

| SOLUTION | | % DISAPPEARANCE OF IAA | |
|---|---|---|---|
| γ/CC. IAA | γ/CC. RIBOFLAVIN | DARK | LIGHT |
| 25 | 0 | 0 | 0 |
| 25 | 0.01 | 0 | 16.0 |
| 25 | 0.1 | 0 | 22.4 |
| 25 | 1 | 0 | 98.8 |
| 25 | 10 | 0 | 98.4 |

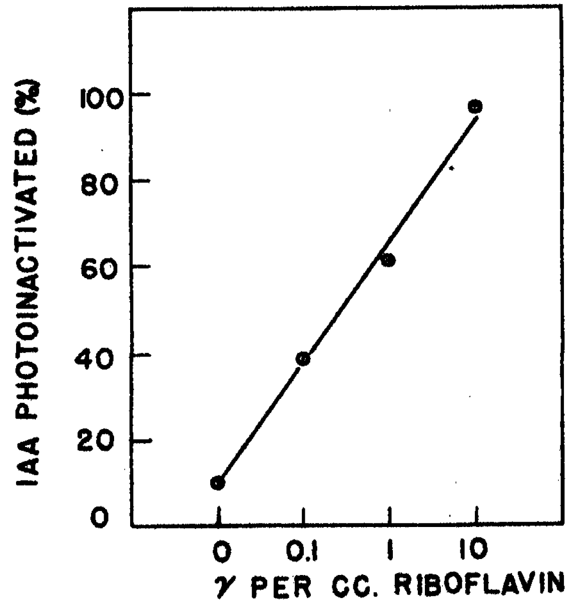


FIGURE 2

The effect of riboflavin concentration on the rate of photoinactivation of IAA. Solution contained 25 γ/cc. IAA; duration of reaction was 40 minutes.

If the log of the residual concentration of IAA is plotted against time, a straight line is obtained, indicating that the reaction is first-order. The reaction rate constant, $K$, as determined by multiplying the slope of this line by 2.303, was $1.2 \times 10^{-4}$ reciprocal seconds. This value checked closely with other determinations of $K$ made by means of the formula

$$K = \frac{2.303}{t} \log \frac{C_o}{C}$$

where $C_o$ = original concentration and $C$ = concentration at time $t$.

To determine more precisely the relation between rate of the reaction and concentration of riboflavin, a series was set up similar to that in table 1, but aliquots for colorimetric analysis were taken after 40 minutes. As shown in figure 2, there is approximately a straight line relationship between the log of the riboflavin concentration and the rate of the reaction. This indicates that the reaction will be strictly first-order with respect to IAA only in the presence of an excess of riboflavin.

The effect of temperature and of pH on the reaction rate are shown in tables 2 and 3. Raising the temperature from 3°C. to 23°C. results in a 50% increase in the rate of the reaction, and further elevation of the temperature to 48°C. has a proportionately smaller effect. The fact that the reaction proceeds best at the lower pH values would indicate that the undissociated molecule reacts more rapidly than does the indoleacetate ion. The pK of indoleacetic acid is 4.75 (D. Bonner[1]).

TABLE 2

THE EFFECT OF TEMPERATURE ON THE RATE OF PHOTO-INACTIVATION OF IAA. ALL SOLUTIONS HAVE 25 $\gamma$/CC. IAA AND 1 $\gamma$/CC. RIBOFLAVIN

| TIME AFTER START OF EXPERIMENT, MIN. | % OF IAA PHOTO-INACTIVATED 3°C. | 23°C. | 48°C. |
|---|---|---|---|
| 15 | 18.0 | 28.8 | 32.0 |
| 30 | 26.8 | 38.9 | 51.9 |
| 45 | 38.4 | 55.3 | 60.0 |
| 60 | 46.8 | 65.6 | 70.4 |

TABLE 3

THE EFFECT OF pH ON THE RATE OF PHOTO-INACTIVATION OF IAA. ALL SOLUTIONS HAVE 25 $\gamma$/CC. IAA, 1 $\gamma$/CC. RIBOFLAVIN, AND 0.01 $M$ $KH_2PO_4$-$Na_2HPO_4$ BUFFER

| pH | % OF IAA PHOTO-INACTIVATED AFTER 40 MINUTES |
|---|---|
| 4.5 | 62.4 |
| 5.6 | 60.7 |
| 6.5 | 56.8 |
| 7.0 | 50.4 |
| 7.4 | 44.7 |
| 8.0 | 40.0 |
| 9.2 | 32.0 |

To determine whether the photo-inactivation requires oxygen, duplicate flasks containing 25 $\gamma$/cc. IAA and 1 $\gamma$/cc. riboflavin were prepared in the dark room. Through one flask a swift stream of argon gas was bubbled for 30 minutes and the flask was then sealed. The other flask was kept as an aerobic control. At zero time, both flasks were brought into the light.

It was noted that the riboflavin in the anaerobic flask became decolorized in about 20 minutes, whereas that in the aerobic flask remained bright yellow. At the end of 1 hour, aliquots were removed from each of the flasks and assayed for residual IAA as before. Whereas 95.5% of the IAA had been photo-inactivated in the aerobic flask, only 11.6% had been photo-inactivated in the anaerobic flask. Admission of oxygen to the latter flask resulted in the rapid disappearance of IAA. This indicates that the reaction may be represented as follows:

$$\mathrm{Rbf} + h\nu \rightarrow \mathrm{Rbf}^*$$
$$\mathrm{IAA} + \mathrm{Rbf}^* \rightarrow \text{oxidation product} + \mathrm{Rbf}\cdot\mathrm{H_2}$$
$$\mathrm{Rbf}\cdot\mathrm{H_2} + {}^1/{}_2\mathrm{O_2} \rightarrow \mathrm{Rbf} + \mathrm{H_2O}$$

The activated riboflavin ($\mathrm{Rbf}^*$) therefore acts as a hydrogen carrier between IAA and oxygen.

In order to elucidate the mechanism of the reaction and the nature of the products formed, experiments were conducted in a Warburg apparatus equipped with a battery of 60-watt bulbs seated under the reaction vessels. Into control cups were placed 1 cc. of either IAA (500 $\gamma$/cc.) or riboflavin (20 $\gamma$/cc.); experimental cups contained both the IAA and riboflavin mixed after the cocks were closed. Vessels with and without KOH in the center well were employed to determine both $O_2$ absorbed and possible $CO_2$ evolved in the reaction. At the conclusion of the experiment, a 1-cc. aliquot of the reaction mixture was removed for the determination of residual IAA, and therefore of the quantity of IAA photoöxidized. Typical data are presented in table 4. They seem clearly to indicate the release of 1 molecule of $CO_2$ per molecule IAA oxidized, and the absorption of about 1 molecule of $O_2$ per molecule oxidized, though oxygen absorption values averaged about 12% higher than this value. A balanced over-all reaction may therefore be written:

$$\mathrm{IAA} + \mathrm{O_2} \xrightarrow[\text{light}]{\text{riboflavin}} \begin{matrix}\text{oxidation}\\ \text{product}\end{matrix} + \mathrm{CO_2}$$

TABLE 4

GAS EXCHANGE DURING PHOTOÖXIDATION OF IAA

| | | $\mu$M $O_2$ CONSUMED | | $\mu$M $CO_2$ EVOLVED | |
|---|---|---|---|---|---|
| EXPT. NO. | $\mu$M IAA OXIDIZED | TOTAL | PER $\mu$M IAA OXIDIZED | TOTAL | PER $\mu$M IAA OXIDIZED |
| P-308 | 2.76 | 3.16 | 1.15 | .. | .. |
| | 2.47 | 2.91 | 1.18 | .. | .. |
| P-314 | 2.73 | 2.58 | 0.95 | 2.78 | 1.02 |
| P-315 | 2.77 | 3.23 | 1.17 | 2.94 | 1.06 |
| | 2.77 | 3.24 | 1.17 | 3.03 | 1.09 |

The nature of the oxidation product is still under investigation and will be reported separately.

*Specificity of the Reactants.*—It became of interest to determine whether compounds related to riboflavin and to indoleacetic acid could react in the same way. Samples of lumichrome and lumiflavin (originally obtained from Dr. J. W. Foster) were made available to the author through the kindness of Dr. H. K. Mitchell. As shown by the data of table 5, lumiflavin was completely incapable of sensitizing the photoöxidation of IAA, whereas lumichrome reacted at a much slower rate than did riboflavin.

TABLE 5

COMPARATIVE EFFECTIVENESS OF RIBOFLAVIN, LUMICHROME AND LUMIFLAVIN IN SENSITIZING THE PHOTOÖXIDATION OF IAA. TEMPERATURE, 21°C.

| SOLUTION | % OF IAA DESTROYED AFTER 1 HR. | 2 HRS. | 17 HRS. |
|---|---|---|---|
| 25 $\gamma$/cc. IAA | 0 | 0 | 0 |
| 25 $\gamma$/cc. IAA + 1 $\gamma$/cc. riboflavin | 70 | 92 | 96 |
| 25 $\gamma$/cc. IAA + 1 $\gamma$/cc. lumichrome | 12 | 16 | 90 |
| 25 $\gamma$/cc. IAA + 1 $\gamma$/cc. lumiflavin | 0 | 0 | 0 |

To determine whether other compounds containing an indole ring could be photoöxidized in the presence of riboflavin, $10^{-4}$ $M$ solutions of such compounds were made up and sensitized by the addition of 10 $\gamma$/cc. riboflavin. At zero time, a Hopkins-Cole color was obtained by mixing 3 cc. of the solution with 1 cc. glyoxylic acid reagent, 0.2 cc. 0.01 $M$ $CuSO_4$ and 5 cc. concentrated $H_2SO_4$. At intervals, aliquots were removed to determine whether the intensity of the Hopkins-Cole color had decreased. Such decrease with respect to the zero time color for each compound was taken as evidence of its photochemical alteration. All indole compounds tested were found to react in the same way as does IAA, although at considerably lower rates. A list of such compounds is found in table 6.

TABLE 6

COMPOUNDS CAPABLE OF BEING PHOTOCHEMICALLY ALTERED IN THE PRESENCE OF 10 $\gamma$/cc. RIBOFLAVIN. ALL SOLUTIONS TESTED WERE $10^{-4}$ $M$

| | |
|---|---|
| Indole | Indole aldehyde |
| Skatole | Tryptamine |
| Indoleacetic acid | Tryptophane |
| Indolebutyric acid | 5-Methyl tryptophane |

*Possible Biological Significance of the Reaction.*—It has been reported by several investigators (van Overbeek,[16] Burkholder and Johnston,[3] Koningsberger and Verkaaik,[9] Stewart and Went,[13] Oppenoorth,[11] Gustafson[6]) that tissue exposed to strong light has a lower auxin content than does unilluminated tissue. If this auxin is indoleacetic acid, then the reaction described in this paper may proceed, destroying either IAA, its precursor, tryptophane (Wildman, *et al.*[18]), or both. Whereas it has been reported (van Overbeek[15]) that only auxin *a* and *b* may produce differential growth effects in light and dark, this reaction provides a mechanism whereby indoleacetic acid may also yield such differential effects, as previously reported by Galston and Hand.[5]

Since the absorption maximum of riboflavin in the visible portion of the spectrum (Ca 460 m$\mu$) is quite close to the most effective wave-length in the action spectrum for phototropism, it is also possible that riboflavin is a receptor pigment in the phototropic response of plants. Recent demonstrations (Galston,[4] Bonner and Thurlow,[2] Leopold and Thimann[10]) that there may be a relation between auxins and photoperiodic response of plants may also be interpreted by means of such a reaction, although the reported action spectrum for photoperiodism (Parker, *et al.*[12]) seems to preclude the participation of riboflavin. These and related questions are at present under investigation in this laboratory.

*Summary.*—In *in vitro* studies, it has been found that riboflavin can sensitize the photoöxidation of various indole-containing compounds, including the plant growth hormone indoleacetic acid. Under the conditions studied, the reaction goes to completion in 1 hour. Kinetic studies indicate that it is first-order.

The reaction requires oxygen. Warburg manometric studies indicate that approximately 1 mol. $O_2$ is absorbed and 1 mol. $CO_2$ is liberated per mol. indoleacetic acid oxidized. The nature of the oxidation product is as yet unknown.

The reaction proceeds best at the lower pH values tested, suggesting that the undissociated IAA molecule reacts more quickly than the indoleacetate ion. The reaction rate is somewhat accelerated by an increase in temperature. Riboflavin analogs work more slowly than riboflavin, or not at all. All indole-containing compounds tested reacted.

It is suggested that riboflavin-sensitized photo-inactivation of auxin may be an important phenomenon in the normal physiology of the plant.

---

[*] The author wishes to express his appreciation to Margery E. Hand and Rosamond S. Baker for their expert technical aid during the course of this investigation.

[1] Bonner, D., *Bot. Gaz.*, **100**, 200–214 (1938).

[2] Bonner, J., and Thurlow, J. F., *Ibid.* (in press).

[3] Burkholder, P. R., and Johnston, E. S., *Smithsonian Miscell. Coll.*, **95**, 1–14 (1947).

[4] Galston, A. W., *Am. J. Bot.*, **34**, 356–360 (1947).

[5] Galston, A. W., and Hand, M. E., *Ibid.*, **36** (in press) (1949).

[6] Gustafson, F. G., *Pl. Physiol.*, **21**, 49–62 (1946).

[7] Haagen-Smit, A. J., Dandliker, W. B., Wittwer, S. H., and Murneek, A. E., *Am. J. Bot.*, **33**, 118–120 (1946).

[8] Haagen-Smit, A. J., Leech, W. D., and Bergren, W. R., *Ibid.*, **29**, 500–506 (1942).

[9] Koningsberger, V. J., and Verkaaik, B., *Rec. Trav. Bot. Néerl.*, **35**, 1–13 (1938).

[10] Leopold, A. C., and Thimann, K. V., *Am. J. Bot.* (in press).

[11] Oppenoorth, W. F. F., Jr., *Rec. Trav. Bot. Néerl.*, **38**, 287–372 (1941).

[12] Parker, M. W., Hendricks, S. B., Borthwick, H. A., and Scully, N. J., *Bot. Gaz.*, **108**, 1–26 (1946).

[13] Stewart, W. S., and Went, F. W., *Ibid.*, 101, 706–714 (1940).
[14] Tang, Y. W., and Bonner, J., *Arch. Biochem.*, 13, 11–25 (1947).
[15] van Overbeek, J., PROC. NAT. ACAD. SCI., 22, 188–190 (1936).
[16] van Overbeek, J., *J. Gen. Physiol.*, 20, 283–309 (1936).
[17] Went, F. W., and Thimann, K. V., *Phytohormones*, Macmillan, 1937.
[18] Wildman, S. G., Ferri, M. G., and Bonner, J., *Arch. Biochem.*, 13, 131–144 (1947).

---

# *AGGREGATION PHENOMENA IN EGG ALBUMIN SOLUTIONS AS DETERMINED BY LIGHT SCATTERING MEASUREMENTS**

BY M. BIER AND F. F. NORD

DEPARTMENT OF ORGANIC CHEMISTRY,** FORDHAM UNIVERSITY, NEW YORK 58, N. Y.

Communicated by F. G. Keyes, November 23, 1948

The various ultracentrifugal measurements reported for the molecular weight of egg albumin are not in complete agreement, as values between 34,500 and 46,000 have been obtained.[1,2] Investigating the behavior of this protein, Sjögren and Svedberg[3] have reported also that, within the pH range 4 to 9, egg albumin is stable and monodisperse.

During the course of extensive studies on the effect of freezing on various proteins and enzymes, it was found, however, that freezing causes, solely because of its physical influence, a change in the state of aggregation of egg albumin, inducing an aggregation at higher concentrations and a disaggregation at lower concentrations of the protein particles in solution.[4] These results were recently confirmed by Neduzhii[5] and others. Furthermore, experiments of Putzeys and Brosteaux,[6] and of Heller and Klevens,[7] extended these results by noting that egg albumin also undergoes aggregation at normal temperatures within the above mentioned pH range of alleged stability.

The causes influencing the state of aggregation of this protein under the above conditions were not yet investigated, and hence we studied the factors governing the aggregation occurring at normal temperature. To study the influence of concentration and time, we adopted the light-scattering method, which is very suitable for the study of aggregation phenomena,[7] the rapidity of the single measurements offering a notable advantage over the determination of particle weights by ultracentrifugation. This is to our knowledge the first attempt of a study of the mechanism of protein aggregation and denaturation by a continuous direct measurement of changes of particle weights.

Regarding the kinetics of aggregation, Oster[8] has pointed out that, in a system of polymerizing or aggregating particles, the light scattering increases with the progress of the reaction. Namely according to Rayleigh,[9] the intensity of scattered light $\tau$, called also turbidity, and expressed in terms of the extinction coefficient, is proportional to the number $N$ of particles per unit volume, and also to the square of the volume of each particle:

$$\tau = ANV^2 \tag{1}$$

Although, because of aggregation, the number of particles is decreasing, the mean volume of each particle is increasing in the same proportion. In a polydisperse system the intensity of the scattered light is the sum of the contributions from each of the components of the mixture.

Equation (1) is limited to a system of independent, dielectric, spherical particles of small diameter as compared with the wave-length of the light. Putzeys and Brosteaux[6] have already established the fact that no correction for the turbidity due to depolarization of the scattered light is required in the case of egg albumin, owing to the symmetry of its particles, whereas the radius, $r = 27.2$ Å, of these particles in the hydrated state, as calculated by Bull,[10] is certainly small enough as compared with the applied wave-length ($\lambda_{\text{air}} = 4358$ Å).

For calculating the particle weight $M$ from the turbidity measurements, Debye's[11] formula

$$\frac{Hc}{\tau} = \frac{1}{M} + 2\,BC \tag{2}$$

was applied, where $c$ is the concentration, $B$[†] is a constant for a determined system, and $H$ is a function of the wave-length and of the refractive indices of the solvent and the solution. For calculating this constant we used the specific refractive increment of egg albumin reported by Barker.[12] For computing the average particle weight at various states of aggregation, it was assumed that the specific refractive increments as well as the coefficient $B$ are not affected by the state of aggregation. For highly aggregated particles, a correction for the asymmetry of the light scattering would probably be necessary; this, however, would only further raise the reported values.

*Experimental.*—To determine the intensity of the scattered light we modified the photoelectric Tyndallometer of the type described by P. P. Debye.[13] The instrument[‡] was calibrated with a sample of polystyrene, the absolute turbidity of which was determined in Debye's laboratory.[§]

The egg albumin was prepared according to the method of La Rosa,[14] recrystallized twice, and dialyzed against distilled water for 3 days. The preparation of the protein was carried out at 4°C., the resulting 6% solution being stored at this same temperature until use.

For each series of measurements the stock solution was centrifuged, filtered in a specially designed sintered glass filter of fine porosity, then transferred into several rectangular glass cells used in the Tyndallometer and there diluted to the various desired concentrations with distilled water.

The cells were then placed in a thermostat and at regular intervals the turbidity of the solutions measured, first allowing sufficient time for the solutions to reach the temperature of the bath. The pH of the solution was determined with a Cambridge meter and found to be 4.18–4.20, independent of the dilution of the protein solution (no buffers used), remaining unchanged throughout the time of each experiment.

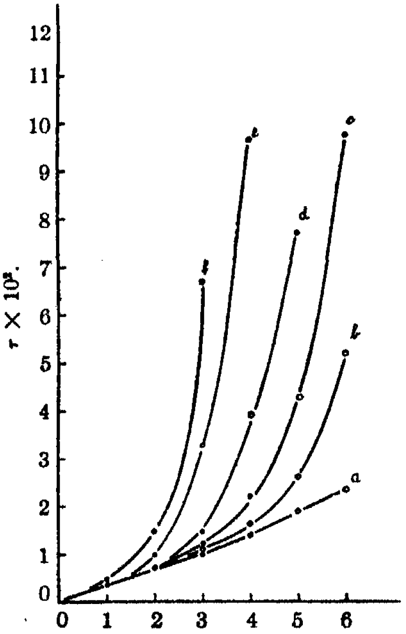


FIGURE 1

Change of turbidity with time as a function of concentration. (*a*) 3 hrs. at 25°C.; (*b*) 20 hrs. at 25°C.; (*c*) 2 hrs. at 35°C.; (*d*) 5 hrs. at 35°C.; (*e*) 13 hrs. at 35°C.; (*f*) 24 hrs. at 35°C.

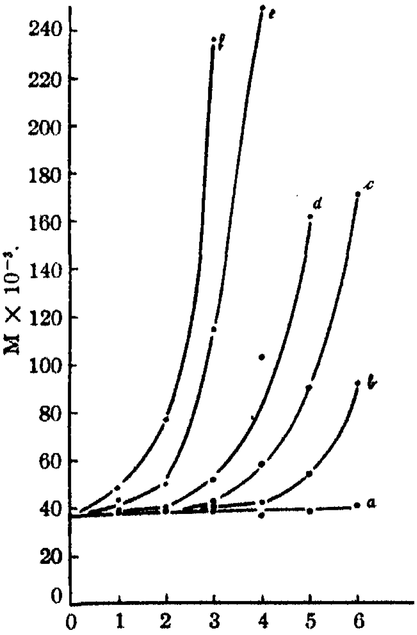


FIGURE 2

Change of average particle weight with time as a function of concentration. For explanation of curves see legend for figure 1.

*Results.*—Turbidity-concentration curves for the albumin solutions at varying time intervals are shown in figure 1, the concentrations ranging from 1 to 6%. The temperature was initially kept at 25°C., and then raised to 35°C., as indicated in the legend. Due to the fact that the coefficient $B$ of equation (2) was found to be approximately zero, there is a nearly linear relationship between the initial turbidities (curve 1*a*) of the various solutions and their concentrations. The increase in turbidity with time, which is characteristic of the more concentrated solutions, is due to aggregation.

At the temperature of 25°C. only the 5 and 6% solutions show a rapid aggregation, whereas the solutions of lower concentration are relatively stable. As we raised the temperature to 35°C. a rapid aggregation occurred also with the more dilute solutions, namely at 2, 3 and 4%, whereas the turbidity of the 1% solution still remained nearly constant. Thus, a very marked influence of concentration on the rate of aggregation is evident, and may be immediately recognized from the shape of the curves. The aggregation proceeds until a visible cloudiness appears, when the turbidity readings were stopped. This is the reason why the curves 1*d*, *e* and *f* do not show the values for the more concentrated solutions.

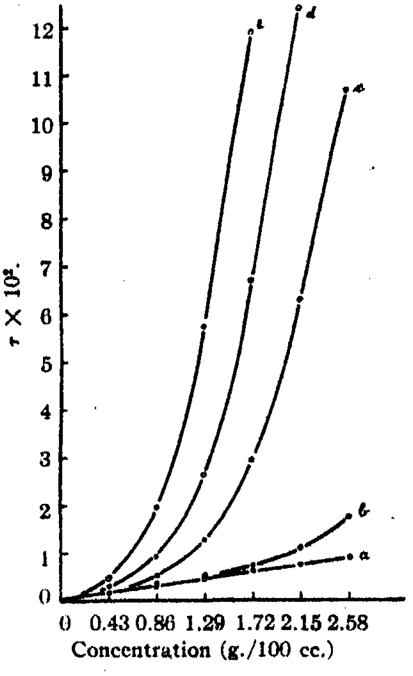


FIGURE 3

Change of turbidity with time as a function of concentration. (*a*) 1 hr. at 20°C.; (*b*) 2 hrs. at 40°C.; (*c*) 14 hrs. at 40°C.; (*d*) 24 hrs. at 40°C.; (*e*) 48 hrs. at 40°C.

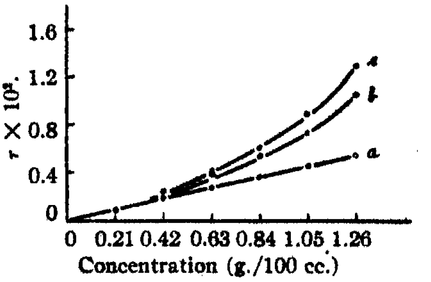


FIGURE 4

Change of turbidity with time as a function of concentration. (*a*) 1 hr. at 25°C.; (*b*) 6 days at 25°C.; (*c*) 1 day at 30°C.

The same results in terms of average particle weight of the egg albumin, calculated on the basis of the above-mentioned assumptions, are presented in figure 2. Each curve clearly indicates the relation between the concentration and aggregation. By extrapolating to infinite dilution, a relatively low value of 37,000 for the molecular weight of egg albumen was obtained, which, however, is within the range of the values recorded in the literature.

In figures 3 and 4 is shown the influence of concentration and time on the turbidity of more dilute solutions: namely, 0.43–2.58% and 0.21–1.26%, respectively. Again an initial linear relationship between turbidity and concentration was found. On standing, the turbidity increases indicating that aggregation has occurred, the rate being of the same over-all dependence on temperature and concentration as above. Presentation of the results in terms of particle weight is omitted, as the curves are similar to those recorded in figure 2.

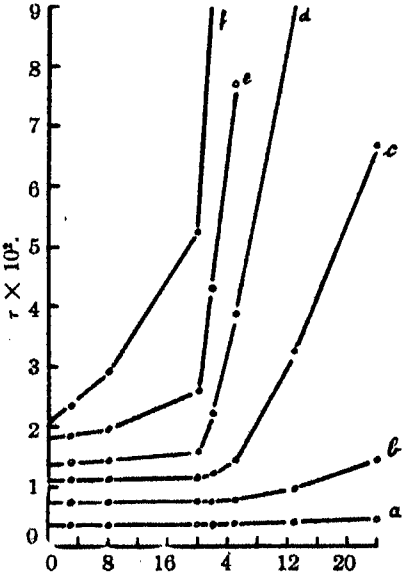


FIGURE 5

Change of turbidity as a function of time. During the first 20 hrs. the temperature was kept at 25°C., then raised to 35°C. (*a*) 1% solution; (*b*) 2% solution; (*c*) 3% solution; (*d*) 4% solution; (*e*) 5% solution; (*f*) 6% solution.

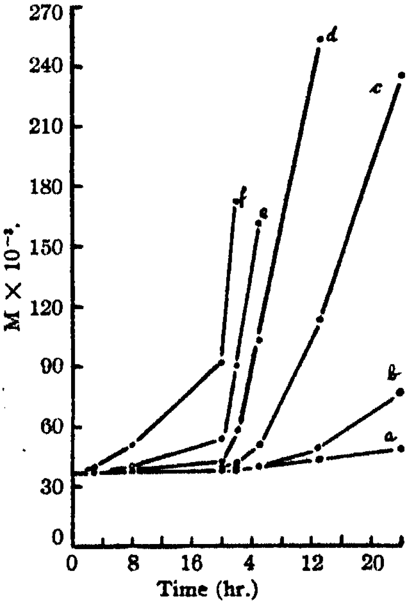


FIGURE 6

Change of average particle weight as a function of time. For explanation of curves see legend for figure 5.

In figure 5 the turbidity values presented in figure 1 are plotted against the time of observation. The turbidity increases nearly linearly with time when the temperature is kept constant. The strong inflection of the curves is caused by the change of temperature from 25 to 35°C. Similar is the relationship of the increase in particle weight of the egg albumin with time, presented in figure 6. The same linear dependence of aggregation on time was observed in all the other experiments, particularly in that shown in figure 4, where the very slow aggregation, due to low concentration and low temperature, was measured over a period of 6 days.

*Discussion.*—The rate of aggregation of dialyzed crystalline egg albumin was found to be particularly dependent upon the temperature and the concentration of the protein solution. However, thus far, no simple mathematical relationship could be found between the concentration and the rate of aggregation. The temperature coefficient $Q_{10}$, defined as the ratio of the reaction rate constants $k$ at the temperatures $T + 10$ to that at $T$, i.e., $k_{T\,+\,10}/k_T$ may be calculated from the relation deduced by Oster,[8] where the turbidity, $\tau$, at any time, $t$, can be expressed by the relation

$$\tau_T = AN_0V_0^2(1 + k_Tt) \qquad (3)$$

assuming that the aggregation is a second-order reaction analogous to the linear polymerization. From equation (3), by derivation, we easily obtain

$$Q_{10} = \frac{k_{T'+10}}{k_{T'}} = \frac{(\partial\tau/\partial t)_{T\,=\,T'\,+\,10}}{(\partial\tau/\partial_t)_{T\,=\,T'}} \qquad (4)$$

Whereas for ordinary chemical reactions $Q_{10}$ is 2–3, temperature coefficients higher than 600 have been reported for protein denaturations.[15] The temperature coefficient of our aggregation reaction, while assuming intermediate values between the above expressed limits, was not constant but continued to increase with rise in temperature. Further experiments over a wider temperature range will be reported later.

The difference in the temperature constants of the aggregation reaction just mentioned and the heat denaturation of egg albumin indicates that the two phenomena are not identical. Supporting evidence for this postulation is that our solutions during the aggregation do not show any measurable increase in free sulfhydryl groups, and do not seem to have their ability to crystallize impaired. Both phenomena are characteristic for the heat denaturation. The phenomenon seems also to differ from the denaturation of egg albumin by acids, as this occurs below pH 4, with larger aggregates appearing only below pH 3.[8]

We have, therefore, to consider egg albumin as being a very labile system, upon the state of aggregation of which a small change in environment, such as concentration or temperature, has a profound effect. The influence of other factors, such as pH and ionic strength, and the particle size distribution of the aggregates, will be discussed in future communications.

*Summary.*—The particle weight of crystalline egg albumin was determined by the light-scattering method. Aggregation phenomena were observed at pH 4.20, i.e., within the usually accepted pH range of stability. Whereas the turbidity increases linearly with time, the rate of this increase,

coinciding with the progress of aggregation, was found to be particularly dependent on the concentration and the temperature of the solution. During the aggregation no increase in free sulfhydryl groups could be detected, and the protein did not seem to have its ability to crystallize impaired.

* This work had the support of grants of the BACHE FUND of the NATIONAL ACADEMY OF SCIENCES and the Office of Naval Research.

** Communication No. 168.

† Edsall[16] has recently discussed the influence of the pH and ionic strength on the coefficient *B* of the proteins.

‡ Built by Mr. W. H. Baker, Elmhurst, L. I., N. Y.

§ The sample was obtained through the courtesy of Mr. T. J. Deszczynski of the Department of Chemistry, Columbia University, New York, N. Y.

[1] Svedberg, T., and Pedersen, K. O., *The Ultracentrifuge*, Oxford (1940), p. 382.

[2] Svedberg, T., *Nature*, **128**, 999 (1931).

[3] Sjögren, B., and Svedberg, T., *J. Am. Chem. Soc.*, **52**, 5187 (1930).

[4] Nord, F. F., Ranke-Abonyi, O. M. v., and Weiss, G., *Ber.*, **65**, 1148 (1932); Ranke-Abonyi, O. M. v., and Nord, F. F., *Kolloid-Z.*, **58**, 198 (1932); Weiss, G., and Nord, F. F., *Z. physik. Chem.*, **166A**, 1 (1933); Lange, F. E. M., and Nord, F. F., *Biochem. Z.*, **278**, 173 (1935); Holzapfel, L., *Kolloid-Z.*, **85**, 272 (1938); Bull, H. B., *Z physik. Chem.*, **161A**, 192 (1932).

[5] Neduzhii, A. A., *J. Applied Chem.* (*U.S.S.R.*), **19**, 535 (1946).

[6] Putzeys, P., and Brosteaux, J., *Trans. Faraday Soc.*, **31**, 1314 (1935).

[7] Heller, W., and Klevens, H. B., *Phys. Rev.*, **67**, 61 (1945).

[8] Oster, G., *J. Colloid Sci.*, **2**, 291 (1947).

[9] Lord Rayleigh, *Phil. Mag.*, **47**, 375 (1899).

[10] Bull, H. B., *Advances in Enzymol.*, **1**, 1 (1941).

[11] Debye, P., *J. Phys. & Colloid Chem.*, **51**, 18 (1947).

[12] Barker, H. A., *J. Biol. Chem.*, **104**, 667 (1934).

[13] Debye, P. P., *J. Applied Phys.*, **17**, 392 (1946).

[14] La Rosa, W., *Chemist-Analyst*, **16**, 3 (1927).

[15] Neurath, H., Greenstein, J. P., Putnam, F. W., and Erickson, J. O., *Chem. Revs.*, **34**, 157 (1944).

[16] Lontie, R., Morrison, P. R., Edelhoch, H., and Edsall, J. T. Paper presented at the 114th Meeting of the A. C. S., Washington, 1948.

---

# *UNUSUAL GENE-CONTROLLED COMBINATIONS OF CARBOHYDRATE FERMENTATIONS IN YEAST HYBRIDS**

BY CARL C. LINDEGREN AND GERTRUDE LINDEGREN

SOUTHERN ILLINOIS UNIVERSITY, CARBONDALE

Communicated by C. F. Cori, November 4, 1948

The ability of Saccharomyces to ferment different sugars is gene-controlled and we have built up breeding stocks which are differentiated by their abilities to ferment galactose, melibiose, maltose, sucrose, raffinose and alpha methyl glucoside (1944, 1947, 1948). Many combinations of gene-controlled fermentative abilities which have not been hitherto described have been produced by hybridization. The criterion of gene-control is the fact that when a large number of tetrads is isolated from a hybrid, two cultures from each tetrad are fermenters and two are non-fermenters.

*The Fermentation Test.*—The fermentation test is performed in a 12- × 100-mm. fermentation tube containing 3 cc. of a 4 per cent concentration of the carbohydrate and an inverted insert, 7 × 50 mm. Each tube receives 2 cc. of the following medium, without added sugar, is plugged with cotton, and sterilized with the inverted insert in place.

| COMPONENTS | AMOUNT PER LITER |
|---|---|
| Peptone | 3.5 g. |
| $MgSO_4$ | 1 g. |
| $KH_2PO_4$ | 2 g. |
| Liquid Yeast Extract, 50% dry matter | 2 cc. |
| Inositol | 10,000 µg. |
| Thiamin | 200 µg. |
| Pyridoxine | 200 µg. |
| Pantothenate | 200 µg. |
| Nicotinic | 200 µg. |
| Para-amino-benzoic acid | 50 µg. |
| Biotin | 2 µg. |
| Adenine | 2,000 µg. |
| Methionine | 2,000 µg. |

One cubic centimeter of a Seitz-filtered 12% solution of Pfanstiehl carbohydrate is then added to make a final concentration of 4%. The tube is inoculated with a large loop of organisms from an agar slant and fermentation generally occurs within 24 to 48 hrs. at 30°C. and results in filling the inverted insert with gas. In addition to the gas production, an abundant growth occurs. Occasionally growth occurs without gas production, but this is rare.

The yeasts described are all able to ferment glucose and fructose (and mannose), and any hydrolysis which splits off one of these hexoses results in a positive fermentation test. The conditions in the tube are relatively anaerobic and whether an aerated culture might ferment differently has not been determined.

The fermentations are adaptive except for glucose and fructose. A positive test may involve "mutation" and selection of a mutant from an originally non-fermenting stock, since considerable growth occurs before fermentation begins, or adaptation may occur without mutation. No attempt was made to distinguish between these possibilities, for single character difference is all that one seeks to demonstrate. The tubes are observed for several weeks, but active fermenters usually fill the insert with gas within 36 hours. There are occasional slow fermenters that may require from 5 to 10 to 15 days before fermentation occurs.

*Galactose-Melibiose Combinations.*—The fermentation of galactose and the hydrolysis of melibiose are controlled by two independently segregating genes. This means that four kinds of progeny are produced when a gamete capable of fermenting both galactose and melibiose (*G ME*) is mated to one incapable of fermenting either sugar (*g me*).

PRODUCTION OF $CO_2$ WHEN DIFFERENT SUGARS ARE USED IN THE FERMENTATION TEST

| GENOTYPE | GALACTOSE | MELIBIOSE |
|---|---|---|
| *G ME* | + | + |
| *G me* | + | – |
| *g ME* | – | + |
| *g me* | – | – |

Over 1000 asci have been analyzed in which these two alleles segregate independently of each other; the gene controlling the ability to ferment galactose does not interact with the gene controlling the ability to hydrolyze melibiose. The existence of a yeast capable of fermenting melibiose, but incapable of fermenting galactose, raised an interesting point, since melibiose is a disaccharide containing glucose and galactose. Fermentation tubes containing galactose or melibiose were inoculated with the four different genotypes and, after ten days, each tube was tested for the presence of residual galactose by the mucic acid test. The following table shows the results:

TESTS FOR THE PRESENCE OF GALACTOSE BY THE *Mucic Acid Test* IN FERMENTATION TUBES AFTER 10 DAYS

| GENOTYPE | GALACTOSE | MELIBIOSE |
|---|---|---|
| *G ME* | – | – |
| *G me* | – | + |
| *g ME* | + | + |
| *g me* | + | + |

Residual melibiose is hydrolyzed by the nitric acid to produce glucose and galactose, and the galactose is converted into mucic acid. These results suggest that the *ME* gene controls the production of a hydrolytic enzyme which breaks melibiose down to glucose and galactose. In the absence of the galactose gene, the glucose is fermented and the galactose is left behind.

*Raffinose.*—Raffinose is a trisaccharide in which fructose, glucose, and galactose are linked together with glucose as the central molecule. The fructose and glucose molecules in raffinose are united by a sucrose linkage, and the ability to ferment one-third raffinose depends on the presence of the sucrose gene, which controls the production of a hydrolytic enzyme that hydrolyzes raffinose into fructose and melibiose. The fructose molecule is fermented and the residual melibiose is left behind.

We have extensive yeast pedigrees in which the *SU/su* genes segregate regularly, producing offspring half of which are fermenters of sucrose and raffinose, and half of which are non-fermenters. We also have stocks distinguished by their ability and inability to ferment melibiose; *ME su* yeasts are completely incapable of fermenting raffinose.

PRODUCTION OF $CO_2$ WHEN DIFFERENT SUGARS ARE USED IN THE FERMENTATION TUBES

| GENOTYPES | MELIBIOSE | SUCROSE | RAFFINOSE |
|---|---|---|---|
| *ME SU* | + | + | + |
| *ME su* | + | – | – |
| *me SU* | – | + | + |
| *me su* | – | – | – |

Unless some evidence is forthcoming on differential permeabilities, these data indicate that (1) the hydrolysis of raffinose depends on the presence of sucrase, which is capable of hydrolyzing fructose from the raffinose molecule, leaving residual melibiose, (2) melibiase is not capable of hydrolyzing melibiose in the raffinose molecule, (3) the complete hydrolysis of raffinose depends on the presence of both melibiase and sucrase, the latter acting first.

*Alpha Glucosides.*—In alpha methyl glucoside and maltose, glucose is linked to the molecule by an alpha glucoside bond. Alpha methyl glucoside is used as an indicator of maltase because of the widely accepted view that they are both hydrolyzed by the same enzyme. We have discovered, however, that different genes control the ability to hydrolyze alpha methyl glucoside and maltose. The gene controlling the hydrolysis of alpha methyl glucoside is segregated independently of that controlling the hydrolysis of maltose and four kinds of yeast genotypes are obtained from heterozygous hybrids: *MG MA*, *Mg ma*, *mg MA* and *mg ma*.

The *MG ma* and *mg MA* genotypes are unexpected since alpha (methyl) glucosidase is supposedly capable of hydrolyzing both maltose and alpha methyl glucoside. It has frequently been shown that the same enzyme is capable of hydrolyzing both substances *in vitro*. The data which we present do not contradict this view; the alternative explanation would require that the membranes of non-fermenters are impermeable to the specific carbohydrate. The possibility that maltose is toxic to the maltose non-fermenters has been tested and excluded by the fact that in mixtures of alpha methyl glucoside and maltose, the alpha methyl glucoside is fermented. *MG ma* yeasts which are actively fermenting alpha methyl glucoside, when introduced into maltose, do not continue to ferment; the *ma* gene is a stable non-fermenter in the *MG ma* type.

The existence of this variety of genotypes suggests either (1) that specifically different enzymes control the hydrolysis of alpha methyl glucoside and maltose or (2) that the genes involved control specifically different permeabilities.

From heterozygous hybrids, we obtain eight kinds of progeny with regard to the fermentation of alpha methyl glucoside, maltose and sucrose. Regularly segregating $SU/su$ alleles can be introduced with the stock and members of each of the above four classes incapable of fermenting sucrose are available.

We are indebted to Drs. S. Hestrin, J. O. Lampen, and Carl F. Cori for valuable criticism and helpful advice.

[*] This work was supported by grants from Anheuser-Busch, Inc., American Cancer Society, and the U. S. Public Health Service.

Leibowitz, J., and Hestrin, S., *Advances in Enzymol.*, **5**, 87–127 (1945).

Lindegren, Carl C., Spiegelman, S., and Lindegren, Gertrude, PROC. NAT. ACAD. SCI., **30**, 346–352 (1944).

Lindegren, Carl C., and Lindegren, Gertrude, *Cold Spring Harbor Symposia Quant. Biol.*, **11**, 115–129 (1947).

Lindegren, Carl C., and Lindegren, Gertrude, *Ann. Missouri Botan. Garden*, **34**, 95–99 (1947).

Lindegren, Carl C., and Gertrude Lindegren (in press).

Dr. Caroline Raut has eliminated the possibility that the MA and MG genes control permeability to maltose and alpha methyl glucoside by showing that the activities of enzyme preparations correspond to the fermentation diagnoses. These facts have led to the conclusion that the MA gene controls the rate of hydrolysis of maltose and has little or no effect on hydrolysis of alpha methyl glucoside and vice versa.

---

# A THEOREM ON CONVEX BODIES OF THE BRUNN-MINKOWSKI TYPE

BY HERBERT BUSEMANN

UNIVERSITY OF SOUTHERN CALIFORNIA, LOS ANGELES

Communicated by S. Lefschetz, October 7, 1948

The present note establishes a theorem on convex bodies in the $n$-dimensional Euclidean space $E^n$ which is a counterpart to the Brunn-Minkowski Theorem. The following formulation of the latter [compare Bonnesen-Fenchel, *Theorie der konvexen Körper*, Berlin, 1934 (quoted as B. F.), pp. 71, 72] will exhibit the analogy:

In $E^n$ let $K$ be a convex body with interior points and $P$ a two-dimensional half-plane bounded by the line $L$. Let the hyperplane normal to $L$ intersect $K$ in the non-empty set $H \cap K$ with $(n - 1)$-dimensional volume $V(H \cap K)$. If $V^{1/(n-1)}(H \cap K)$ is laid off from the point $H \cap L$ on the ray $H \cap P$, then the resulting curve in $P$ is convex (and turns its concavity towards $L$).

The present theorem is concerned with the intersection of $K$ by pencils of non-parallel hyperplanes:

I. *In* $E^n$ *let* $K$ *be a convex body with interior points,* $L$ *an* $(n - 2)$*-dimensional linear space which intersects* $K$*, and* $P$ *a 2-dimensional space normal to* $L$ *at a point* $0$*. Any half-hyperplane* $H$ *bounded by* $L$ *intersects* $K$ *in a non-empty set* $H \cap K$*. If the* $(n - 1)$*-dimensional volume* $V(H \cap K)$ *of* $H \cap K$ *is laid off from* $0$ *on the ray* $H \cap P$ *then the resulting curve* $C$ *is convex* ($C$ contains 0 in its interior if $L$ contains interior points of $K$, otherwise $P$ lies on $C$).

Both theorems remain true, but are trivial if $K$ has no interior points.

Notice the following Corollary of I:

II. *Let* $K$ *be a convex body in* $E^n$ *with* $0$ *as interior point and center. If for any hyperplane* $H$ *through* $0$ *the volume* $V(H \cap K)$ *is laid off from* $0$ *on the normal to* $H$ *at* $0$ *(in both directions), then the resulting surface* $S$ *is convex.*

II follows from I because it suffices to know that the intersection of $S$ with a two-dimensional plane $P$ through 0 is a convex curve $C'$. But $C'$ originates from the $C$ of I by revolving $C$ in $P$ about 0 through $\pi/2$ and then dilating it in the ratio 2: 1.

The result II is decisive for the theory of area in Finsler spaces and more specifically in Minkowski (or finite dimensional Banach) spaces. Its applications will be found in the author's series of articles on "Curvature and Angle in Non-Riemannian Spaces" which is appearing elsewhere.

Theorem I is trivial for $n = 2$, so that $n \geqslant 3$ will be assumed in the proof. It may also be assumed that $L$ contains interior points of $K$, because the other case follows from this one by an obvious limit process. Then $V(H \cap K) > 0$ for every half-hyperplane $H$ through $L$.

Let $\mathfrak{r}$ and $\mathfrak{s}$ be non-opposite rays in $P$ with origin 0, and $\mathfrak{q}$ a ray with origin 0 in the angle with legs $\mathfrak{r}$ and $\mathfrak{s}$. Let the segment connecting the points $r \in \mathfrak{r}$, $s \in \mathfrak{s}$ $(r, s \neq 0)$ intersect $\mathfrak{q}$ at $q$. If $\angle (\mathfrak{r}, \mathfrak{q}) = \alpha'$, $\angle (\mathfrak{q}, \mathfrak{s}) = \beta'$ then the fact that the area of the triangle $ors$ equals the sum of the areas of the triangles $orq$ and $oqs$ yields

$$or\, os \sin (\alpha' + \beta') = or\, oq \sin \alpha' + oq\, os \sin \beta',$$

hence

$$oq = or\, os \sin (\alpha' + \beta')/(or \sin \alpha' + os \sin \beta'). \qquad (1)$$

Let $H_r$, $H_s$, $H_q$ denote the half-hyperplanes bounded by $L$ and through $\mathfrak{r}$, $\mathfrak{s}$, $\mathfrak{q}$, respectively. Then (1) shows that I is equivalent to the statement

$$V(H_q \cap K) \geqslant \frac{\sin (\alpha' + \beta') V(H_r \cap K) V(H_s \cap K)}{\sin \alpha' V(H_r \cap K) + \sin \beta' V(H_s \cap K)} \qquad (2)$$

With the notations $\alpha = \sin \alpha'$, $\beta = \sin \beta'$, $\delta = \sin (\alpha' + \beta')$ and $V(H_q \cap K) = D$, $V(H_r \cap K) = A$, $V(H_s \cap K) = B$, the relation (2) becomes

$$D \geqslant \delta AB/(\alpha A + \beta B). \tag{3}$$

Let $\bar{A}_x(\bar{B}_y)$ denote the set at which the $(n - 2)$-dimensional space parallel to $L$ through the point $r$ with $or = x$ on $\mathfrak{r}$ ($s$ with $os = y$ on $\mathfrak{s}$) intersects $K$. If $A_x(B_y)$ is the $(n - 2)$-dimensional measure of $\bar{A}_x(\bar{B}_y)$, then

$$A = \int A_x dx \qquad B = \int B_y dy.$$

Following the original idea of Brunn for the proof of the Brunn-Minkowski Theorem, we map the interval of $\mathfrak{r}$ for which $\bar{A}_x \neq 0$ on the interval of $\mathfrak{s}$ for which $B_y \neq 0$ through the relations

$$\rho A = \int_0^x A_t dt \qquad \rho B = \int_0^y B_t dt, \qquad 0 \leqslant \rho \leqslant 1 \tag{4}$$

so that

$$\frac{dx}{d\rho} = \frac{A}{A_x} \qquad \frac{dy}{d\rho} = \frac{B}{B_y}, \qquad 0 \leqslant \rho < 1. \tag{5}$$

Since $K$ is convex it contains for every $\rho$ the convex closure $C_\rho$ of $\bar{A}_{x(\rho)}$ and $\bar{B}_{y(\rho)}$. Since $\bar{A}_x$ and $\bar{B}_y$ lie in $(n - 2)$-dimensional spaces parallel to $L$, the set $C_\rho$ lies in the hyperplane $H_\rho$ through these two spaces. $H_\rho$ intersects $H_q$ in an $(n - 2)$-dimensional space parallel to $L$ and $\mathfrak{q}$ itself in a point whose distance $u(\rho)$ from 0 is by (1)

$$u(\rho) = \delta x(\rho)y(\rho)/[\alpha x(\rho) + \beta y(\rho)], \tag{6}$$

hence by (5)

$$\frac{du}{d\rho} = \delta\frac{\alpha x^2y' + \beta y^2x'}{(\alpha x + \beta y)^2} = \delta\frac{\alpha Bx^2B_y^{-1} + \beta Ay^2Ax_x^{-1}}{(\alpha x + \beta y)^2}. \tag{7}$$

The intersection $\bar{C}_{u(\rho)}$ of $C_\rho$ with $H_q$ lies in $\bar{D} = H_q \cap K$ because $\bar{C}_{u(\rho)} = C_\rho \cap H_q \subset K \cap H_\rho \cap H_q = \bar{D} \cap H_\rho$. Therefore, if $C_{u(\rho)}$ is the $(n - 2)$-dimensional measure of $\bar{C}_{u(\rho)}$,

$$D \geqslant \int C_u du = \delta \int_0^1 C_{u(\rho)} \frac{\alpha Bx^2B_y^{-1} + \beta Ay^2A_x^{-1}}{(\alpha x + \beta y)^2} d\rho. \tag{8}$$

In the original triangle $orq$ we find, if $V = \angle\, oqr$, that

$$or:rq = \sin V:\alpha, \qquad os:qs = \sin(\pi - V):\beta,$$

hence

$$rq:qs = or\cdot\alpha:os\cdot\beta. \tag{9}$$

If (9) is applied to the points $q$, $r$, $s$ with $oq = u(\rho)$, $or = x(\rho)$, $os = y(\rho)$ then $rq:qs$ equals the ratio of the distances $d_r$, $d_s$ of $H_\rho \cap H_q$ from $H_\rho \cap H$ and $H_\rho \cap H_s$:

$$d_r : d_s = \alpha x : \beta y. \tag{10}$$

Therefore the Brunn-Minkowski Theorem (compare B. F., p. 88) yields applied to $A_{x(\rho)}$, $B_{y(\rho)}$ and $C_{u(\rho)}$ that

$$C_u^{1/(n-2)} \geqslant \frac{\beta y}{\alpha x + \beta y} A_x^{1/(n-2)} + \frac{\alpha x}{\alpha x + \beta y} B_y^{1/(n-2)} \tag{11}$$

hence by (8) and (11)

$$D \geqslant \delta \int_0^1 [\beta y A_x^{1/(n-2)} + \alpha x B_y^{1/(n-2)}]^{n-2} (\alpha B x^2 B_y^{-1} + \beta A y^2 A_x^{-1})(\alpha x + \beta y)^{-n} d\rho. \tag{12}$$

In order to prove (3) it suffices to show that *the integrand in (12) is for* $0 < \rho < 1$ *not less than* $AB(\alpha A + \beta B)^{-1}$. A similar procedure as in the proof of the Brunn-Minkowski Theorem in B. F., p. 89, will be used. If $\epsilon = (n-1)^{-1}$, then

$$\beta y A_x^{1/(n-2)} + \alpha x B y_y^{1/(n-2)} = y^{1+\epsilon}\beta A^{\epsilon}\left(\frac{A_x}{A^{1-\epsilon}y^{1-\epsilon}}\right)^{1/(n-2)} + x^{1+\epsilon}\alpha B^{\epsilon}\left(\frac{B_y}{B^{1-\epsilon}x^{1-\epsilon}}\right)^{1/(n-2)}$$

$$\frac{\alpha B x^2}{B_y} + \frac{\beta A y^2}{Ax} = y^{1+\epsilon}\beta A^{\epsilon}\frac{A^{1-\epsilon}y^{1-\epsilon}}{A_x} + x^{1+\epsilon}\alpha B^{\epsilon}\frac{B^{1-\epsilon}x^{1-\epsilon}}{B_y}$$

Putting $t = x^{1+\epsilon}\alpha B^{\epsilon}[x^{1+\epsilon}\alpha B^{\epsilon} + y^{1+\epsilon}\beta A^{\epsilon}]^{-1}$ the integrand in (12) may be written as

$$\left[(1-t)\left(\frac{A_x}{A^{1-\epsilon}y^{1-\epsilon}}\right)^{1/(n-2)} + t\left(\frac{B_y}{B^{1-\epsilon}x^{1-\epsilon}}\right)^{1/(n-2)}\right]^{n-2} \cdot \left[(1-t)\frac{A^{1-\epsilon}y^{1-\epsilon}}{A_x} + t\frac{B^{1-\epsilon}x^{1-\epsilon}}{B_y}\right] \cdot \frac{(x^{1+\epsilon}\alpha B^{\epsilon} + y^{1+\epsilon}\beta A^{\epsilon})^{n-1}}{(\alpha x + \beta y)^n}. \tag{13}$$

By Jensen's Inequality (compare Hardy, Littlewood, Polya, *Inequalities*, Cambridge, 1934, Theorem 16, p. 26)

$$[(1-t)a_1^{1/\nu} + ta_2^{1/\nu}]^{\nu} \geqslant [(1-t)a_1^{-1} + ta_2^{-1}]^{-1} \quad \text{for } 0 < t < 1,\ a_i > 0,\ \nu > 0 \tag{14}$$

and the equality sign holds only when $a_1 = a_2$.

Application of (14) to (13) with $\nu = n - 2$ yields that the integrand in (12) is at least

$$\frac{(x^{1+\epsilon}\alpha B^{\epsilon} + y^{1+\epsilon}\beta A^{\epsilon})^{n-1}}{(\alpha x + \beta y)^n} = \left(\frac{\alpha x \cdot x^{\epsilon}B^{\epsilon} + \beta y \cdot y^{\epsilon}A^{\epsilon}}{\alpha x + \beta y}\right)^{n-1} \cdot \frac{1}{\alpha x + \beta y}. \tag{15}$$

Applying (14) again with $\nu = n - 1 = \epsilon^{-1}$ and $t = \beta y(\alpha x + \beta y)^{-1}$ to the first factor on the right in (15) yields that (15), and therefore the integrand in (12), is at least

$$\left(\frac{\alpha x\cdot x^{-1}B^{-1}+\beta y.y^{-1}A^{-1}}{\alpha x+\beta y}\right)^{-1}\cdot\frac{1}{\alpha x+\beta y}=\frac{AB}{\alpha A+\beta B}$$

which proves Theorem I.

Conditions for the equality sign in (2) are important for several applications. Because of the continuity of the functions involved, the equality sign in (2) requires the equality sign in (11) and in the two cases of Jensen's Inequality. Moreover $\bar{C}_{u(\rho)}$ must coincide with $\bar{D} \cap H_\rho$.

The equality sign holds in (11) only if the sets $\bar{A}_{x(\rho)}$ and $\bar{B}_{y(\rho)}$ are homothetic (see B. F., pp. 72 and 88). The condition $a_1 = a_2$ for equality in (14) yields in the two cases

$$A_x/A^{1-\epsilon}y^{1-\epsilon} = B_y/B^{1-\epsilon}x^{1-\epsilon} \text{ and } xB = yA$$

so that $A_x = B_y$.

Therefore $\bar{A}_x$ and $x\bar{B}_y$ are not only homothetic but congruent. The relation $x:A = y:B$ shows then that $\bar{A} = H_r \cap K$ can be transformed into $\bar{B} = H_s \cap K$ by revolving $\bar{A}$ about $L$ through $\alpha' + \beta'$ and dilating it at the ratio $B:A$ in the direction of $\zeta$. If $w'$ is the point of $\bar{B}$ into which the point $w$ of $\bar{A}$ is mapped under this transformation, it is easily seen that the segments connecting $w$ to $w'$ form a convex set $E$ when $w$ tranverses $\bar{A}$, so that $E$ is the convex closure of $\bar{A} \cup \bar{B}$. Clearly the equality sign holds in (2) for $K = E$. Thus we find

III. *The equality sign holds in (2) if and only if $H_r \cap K$ can be transformed into $H_s \cap K$ by a rotation about $L$ and a subsequent dilation in the direction of $\zeta$, and $H_q \cap K$ is the intersection of $H_q$ with the convex closure of $H_r \cap K$ and $H_s \cap K$.*

From III and the fact that $E$ is the union of the segments connecting $w$ and $w'$ we obtain the following addition to II:

IIa. *The surface $S$ in II is strictly convex if $K$ is strictly convex.*

But III shows that the strict convexity of $K$ is not a necessary condition.

# THE *n*-ALITY THEORY OF RINGS

By ALFRED L. FOSTER

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA AT BERKELEY

Communicated by G. C. Evans and read before the Academy, November 17, 1948

1. *Introduction.*—The classic duality-symmetry which is exhibited by Boolean rings (and Boolean algebras), far from being characteristic of

this class, is actually a phenomenon inherent (though usually dormant) in all rings. This ring duality has been presented and variously explored in a series of papers.[1–7]

The present communication is intended as a brief partial résumé dealing with extensions of this duality theory to *n-ality*, and more generally to *K-ality* theories, corresponding to various groups $K$ of admissible "co-ordinate transformations." The specialization $K = C$ = complementation group, of order 2 (see §3), yields the original ring duality, which latter we now designate as *simple*, or *mod K*, to distinguish it from rival theories.

The elevation of the simple theory to the general $K$-level, and a proper refinement thereof, throws new light on certain previous results on the $C$-level, particularly in connection with questions dealing with the (simple) logical-algebra definability of rings. However, the most important current results on the $K$-level center around the new concept, *ring-logic* ($K$). Roughly the latter is defined as a ring which both uniquely (equationally) determines and is uniquely (equationally) determined by its $K$-logical-algebra. On the simple level, for instance, a Boolean ring is a ring-logic ($C$).

We introduce *p-rings*, a natural generalization of Boolean rings (which are coextensive with 2-rings), formulate their *p-ality theory*, and establish that 3-rings form a new class of ring-logics. These 3-ring-logics subsume the familiar 3-valued-logic; moreover, the relationship between the latter, general 3-rings, their logical algebras and the enveloping *tri-ality theorem* on the one hand, forms an exact generalization of the relationship between 2-valued logic (= logic of propositions), general Boolean rings (= 2-rings), their logical algebras and the encompassing duality theory on the other.

2. *Fundamentals of the Simple Duality Theory.*—If $(R, +, \times)$ is a ring (with unit), the concepts of $R$ occur in dual pairs. For example, 0 and 1 are dual elements; $\times, \times'$; $+, +'$; $-, -'$; $*$, are respective dual pairs of operations, the latter being self-dual, where

$$\begin{array}{ll} \left.\begin{aligned} a \times' b &= a + b - (a \times b) \\ a \times b &= a +' b -' (a \times' b) \end{aligned}\right\} & \text{dual ring products} \\ \left.\begin{aligned} a +' b &= a + b - 1 \\ a + b &= a +' b -' 0 \end{aligned}\right\} & \text{dual ring sums} \\ \left.\begin{aligned} a -' b &= a - b + 1 \\ a - b &= a -' b +' 0 \end{aligned}\right\} & \text{dual ring differences} \\ \quad a^* = 1 - a = 0 -' a & \text{(self-dual) ring complement} \end{array} \tag{2.1}$$

Restricted for brevity to these concepts one has[1, 5] the

(SIMPLE) RING DUALITY THEOREM. *If* $P(0, 1;\ \times, \times';\ +, +';\ -, -';\ *)$ *is a true proposition of a ring* $(R, +, \times)$, *then so also is its dual*

$$dl.\,P = P(1, 0;\ \times', \times;\ +', +;\ -', -;\ *)$$

*obtained by replacing each argument by its dual, as indicated, with * left unchanged (self-dual).*

Illustrations of the duality theorem in an arbitrary ring are given by the various dual relations (2.1). Again, by

$$\left.\begin{array}{l}(a \times b)^* = a^* \times' b^* \\ (a \times' b)^* = a^* \times b^*\end{array}\right\} \text{ring "De-Morgan" formulas.} \tag{2.2}$$

$$a^{**} = a, \quad 0^* = 1, \qquad 1^* = 0 \tag{2.3}$$

$$\begin{array}{l}(R, +, \times) \text{ is a ring with 0 as zero element and 1 as unit.}\\ (R, +', \times') \text{ is a ring with 1 as zero element and 0 as unit.}\end{array} \tag{2.4}$$

As additional illustrations, in a Boolean ring (or, more generally, in a Boolean-like ring,[1]) one has

$$\begin{array}{l}a + b = (a \times b^*) \times' (a^* \times b)\\ a +' b = (a \times' b^*) \times (a^* \times' b)\end{array} \tag{2.5}$$

Again, in a field, for instance,

$$a + b: \begin{cases}a + b = a \times' b \times' (a^0 \times b^0) & (a \neq 1, b \neq 1)\\ a + 1 = 1 + a = a \times' a^{10} & (a \neq 0, \neq 1)\\ 1 + 1 = a \times' a^{10} \times' a^{01} & (a \neq 0, \neq 1)\\ 1 + 0 = 0 + 1 = 1 & \end{cases}$$

$$a +' b: \begin{cases}a +' b = a \times b \times (a^1 \times' b^1) & (a \neq 0, b \neq 0)\\ a +' 0 = 0 +' a = a \times a^{01} & (a \neq 1, \neq 0)\\ 0 + 0 = a \times a^{01} \times a^{10} & (a \neq 1, \neq 0)\\ 0 +' 1 = 1 +' 0 = 0 & \end{cases} \tag{2.6}$$

(Here $a^1$ and $a^0$ are the $\times$ and $\times'$ inverses of $a$, respectively.)

If $R$ is a Boolean ring, the dual products $\times$, $\times'$ reduce to the logical product, $\cap$, and logical sum, $\cup$, respectively; * to the Boolean complement, $^-$, and duality and "De-Morgan" theorems to the corresponding Boolean theorems.

With this latter specialization as motivation, in an arbitrary ring the (operationally closed) system $(R, \times, \times', ^*)$ was introduced in reference 1 as the (simple) *logical algebra*, or briefly the *logic* of the ring. This is an illustration of a "mixed" concept; a ring in either of the forms $(R, +, \times)$ or $(R, +', \times')$ is "pure," that is, formulated within a single "coordinate system," but "mixed" in the form $(R, +, \times, +', \times', ^*)$. A (simple) logical concept of a ring is one definable in terms of the logical-algebra of the ring; a (simply) *logically definable ring* is one whose ring sum, $+$, and therefore the entire ring, is a simple logical ($=$ $C$-logical) concept. It has been shown that Boolean rings,[1, 3, 5] Boolean-like rings,[1] fields,[4] and integral domains are examples of $C$-logically definable rings.

*3. General Transformation Theory.*—If $U = \{\ldots, x, \ldots\}$ is a class (with or without structure), $\varphi(x, y, \ldots)$ any multitation of (= operation of one or more arguments in) $U$, and $\rho(x)$ any permutation (= 1-1 monotation) of $U$, with inverse written $\rho^-$, then in the "$\rho$ coördinate system" the "point" $x$ receives the new "coördinate" $\rho(x)$, and the multitation $\varphi$ "becomes" $\varphi_\rho$, where

$$\varphi_\rho = \rho^-(\varphi(\rho(x), \rho(y), \ldots). \tag{3.1}$$

Based on this simple extension of the "transform" concept, the traditional (*a*) transformation and invariant theories (various groups, tensors, matrices, etc.), in all of which the underlying (*b*) "computational" disciplines (analysis, arithmetic, a particular ring—in fact any operational algebra) are regarded as absolute invariants (unchanged by coördinate transformations), may be brought under a single unifying theory in which a given such "computational" discipline may itself profitably be thought of as transforming "cogrediently," or sometimes "contragrediently" with changing "coördinates."

We consider here only rings. If $(R, +, \times)$ is a ring and $K$ a preassigned group of coördinate transformations in (= permutations of) the set $R$, then in the $\rho$ coördinate system the basic ring operations become

$$\begin{aligned} a + b &\rightarrow a +_\rho b = \rho^-(\rho(a) + \rho(b)) \\ a \times b &\rightarrow a \times_\rho b = \rho^-(\rho(a) \times \rho(b)) \end{aligned} \tag{3.2}$$

If $+, +', +'', \ldots$ and $\times, \times', \times'', \ldots$ are the (conjugate) classes $\{+_\rho\}$ and $\{\times_\rho\}$ of all transforms (3.1) of $+$ and of $\times$ by the various $\rho \in K$, then $(R, +, \times)$, $(R, +', \times')$, ..., etc., denote the "same" ring, in different coördinates, precisely as in the (subsumed special) case of tensors, matrices, etc., in different coördinates.

If, in particular, the admissible group $K$ of coördinate transformations is chosen as $C$ (order 2),

$$x^* = 1 - x, \quad x^{**} = x = \text{identity}, \tag{3.3}$$

the $C$-ality ring theory reduces to the original *simple* theory. Further specialized to a Boolean ring, for instance, the duality of logical-sum $\cup$ (= $\times'$) and logical-product $\cap$ (= $\times$) is simply a way of saying that these are *identical* notions, expressed in different coördinates.

For general $K$ the *K-logical algebra* (= *K-logic*) of a ring $(R, +, \times)$ is defined as the (operationally closed) system

$$(R, \times, \times', \times'', \ldots, \xi, \rho, \rho', \rho'', \ldots) \tag{3.4}$$

whose operations are the various monotations $\xi$ (= identity), $\rho, \rho', \rho''$, ... of $K$, and the (conjugate set of bitations) $\times, \times', \times''$, ... of transforms (3.1) of $\times$ by the $\xi, \rho, \rho', \ldots$ of $K$.

A ring is: (*a*) *K-logically definable* if its ring sum, $+$, (and hence the whole ring) is a $K$-logical concept; ($b$) $K$-logically *equationally* definable if its $+$ is compositionally generable by the $K$-logical operations $\times$, $\times'$, $\times''$, $\ldots$, $\xi$, $\rho$, $\rho'$, $\ldots$, or, what is equivalent, if its $+$ satisfies an identity

$$a + b \underset{a,\, b}{\equiv} \varphi(a, b) \tag{3.5}$$

where $\varphi$ is compositionally generable by $a \times b$, $\xi(a)(= a)$, $\rho(a)$, $\rho'(a)\ldots$; ($c$) $K$-logically *fixed* if it is $K$-logically definable, and if no ring $(R, +_1, \times)$ exists (on the same class $R$ and with the same $\times$, but with different ring sum $+_1, \neq +$), which is $K$-logically definable.

On the $C$-level the following results are established:

Theorem I. (*1°*) *A Boolean ring is C-logically fixed and equationally definable.* (*2°*) *A Boolean-like ring is C-logically definable but not in general C-logically fixed.* (*3°*) *A field is C-logically fixed, but not in general C-logically equationally definable.*

We introduce the concept *ring-logic* ($K$) as synonymous with a ring which is both $K$-logically fixed and equationally definable. The above Theorem asserts that a Boolean ring is a ring-logic ($C$), while a field, or a Boolean-like ring is generally not. In a ring-logic ($K$) we capture in generalized form the extreme intimacy between ring and logical algebra which, on the $C$-level, is exemplified by the class of Boolean ring-logics. The generalized parallelism is still closer when the $K$-ality theory which permeates ring and logic is recognized.

*4. p-Rings.*—We arrive at a new class of ring-logics by considering *p-rings*, a natural generalization of Boolean rings. A $p$-ring ($p$ = prime integer) is a commutative ring with unit $(S, +, \times)$, in which for all $a \in S$

$$a^p = a \times a \times \ldots \times a = a \tag{4.1}$$

$$pa = a + a + \ldots + a = 0 \tag{4.2}$$

Thus the concepts 2-ring and Boolean ring coincide; for $p = 2$, (4.2) follows from (4.1), but not for $p > 2$. For given $p$, the simplest $p$-ring is $(F_p, + \times)$, the field of residues mod $p$. As in the special Boolean case, all finite $p$-rings are direct products

$$F_p \times F_p \times F_p \times \ldots \times F_p \tag{4.3}$$

of the prime field $F_p$, and hence have $p^t$ elements. We here omit considerations of general ideal structure.

The "cyclic negation group" $N$ (of order $p$) of a $p$-ring is the cyclic group generated by $x^{\wedge}$,

$$x^{\wedge} = 1 + x \tag{4.4}$$

(For 2-rings $N$ is seen to be identical with $C$, since $x^{\wedge} = 1 + x = 1 - x = x^*$ in a Boolean ring.)

Since we here consider only the above group $N$ of coördinate transformations in a $p$-ring, we may omit reference to it. Thus, $p$-ality $\longleftrightarrow$ $p$-ality $(N)$, etc.

Each $p$-ring has a *$p$-ality theory,* that is, each concept or proposition of the ring is one of a $p$-al set of $p$ such concepts (or is self $p$-al, i.e., the same in all admissible coördinates). For convenience, however, we shall deal with the *tri-ality* case ($p = 3$), from which the form of the general $p$-ality theorem will at once be clear. For $p = 3$ we abbreviate $x^{\wedge\wedge} = x^{\vee}$, and hence

$$\begin{aligned} x^{\wedge\vee} = x^{\vee\wedge} = x^{\wedge\wedge\wedge} &= x^{\vee\vee\vee} = x \\ x^{\wedge\wedge} = x^{\vee},\ x^{\vee\vee} &= x^{\wedge}. \end{aligned} \tag{4.5}$$

From (3.1) and (4.4) the tri-als if the ring product $\times$ are given by:

$$\begin{aligned} a \times' b &= (a^{\wedge} \times b^{\wedge})^{\vee} = a \times b + a + b \\ a \times'' b &= (a^{\vee} \times b^{\vee})^{\wedge} = a \times b + 2(a + b) + 2, \end{aligned} \tag{4.6}$$

and similarly

$$\begin{aligned} a +' b &= (a^{\wedge} + b^{\wedge})^{\vee} = a + b + 1 \\ a +'' b &= (a^{\vee} + b^{\vee})^{\wedge} = a + b + 2 \\ a -' b &= (a^{\wedge} - b^{\wedge})^{\vee} = a - b - 1 \\ a -'' b &= (a^{\vee} - b^{\vee})^{\wedge} = a - b - 2 \end{aligned} \tag{4.7}$$

In a formula a "multiplicative" constant, such as the 2 in $2\,(a + b)$ is an *apparent* (or *removable*) constant,

$$2(a + b) = a + b + a + b$$

An "additive" constant is *real.*

THEOREM II. *Tri-ality Theorem for 3-Rings.*

*Let*

$$P(0, 1, 2;\ \times, \times', \times'';\ +, +', +'';\ -, -', -'';\ ^{\wedge};\ ^{\vee})$$

*be any true proposition in a 3-ring, $S$, where $P$ involves no apparent constants. Then each of the propositions tri-al to $P$,*

$$\begin{aligned} P' &= P(2, 0, 1;\ \times', \times'', \times;\ +', +'', +;\ -', -'', -;\ ^{\wedge};\ ^{\vee}) \\ P'' &= P(1, 2, 0;\ \times'', \times, \times';\ +'', +, +';\ -'', -, -';\ ^{\wedge};\ ^{\vee}) \end{aligned} \tag{4.8}$$

*obtained by (a) leaving each of the operations $^{\wedge}$, $^{\vee}$ unchanged, (b) applying any cyclic permutation "cogrediently" to all other operations, and (c) the "contragredient" (= inverse) permutation to the real constants, is again a true proposition of $S$.*

The "contragredient" element is not apparent in the duality case $p = 2$, since in this special case $x^{\wedge} = x^{\vee} = x^{*}$ is its own inverse. The self-tri-ality of $^{\wedge}$ as well as $^{\vee}$ is similar to the situation for $^{*}$ in the Boolean case.

Of the multitude of illustrations of the tri-ality theorem in any 3-ring we can here mention only two of the simplest.

THEOREM III. *Transformation Theorem (or "De Morgan" Formulas) for 3-Rings. In any 3-ring,*

$$\begin{aligned} a \times' b &= (a^{\wedge} \times b^{\wedge})^{\vee} &&= (a^{\vee} \times'' b^{\vee})^{\wedge} \\ a \times'' b &= (a^{\vee} \times b^{\vee})^{\wedge} &&= (a^{\wedge} \times' b^{\wedge})^{\vee} \\ a \times b &= (a^{\wedge} \times'' b^{\wedge})^{\vee} &&= (a^{\vee} \times' b^{\vee})^{\wedge} \end{aligned} \tag{4.9}$$

These are the equations of "conversion" from each permissible coördinate system to any other such. Similarly the conversion formulas for the tri-al ring sums and differences are given by

THEOREM IV. *In any 3-ring,*

$$\begin{aligned} a +' b &= a + b + 1 &&= a +'' b +'' 0 \\ a +'' b &= a + b + 2 &&= a +' b +' 0 \\ a + b &= a +'' b +'' 2 &&= a +' b +' 1 \\ a -' b &= a - b - 1 &&= a -'' b -'' 0 \\ a -'' b &= a - b - 2 &&= a -' b -' 0 \\ a - b &= a -'' b -'' 2 &&= a -' b -' 1 \end{aligned} \tag{4.10}$$

5. *3-Ring-Logics.*—For 3-rings, always referring to the cyclic negation group $N$ (of order 3), we have established

THEOREM V. *Each 3-ring* $(S, +, \times)$ *is a ring-logic, with* $+$ *logically definable by the equation*

$$a + b = ab^{\wedge} \times' a^{\wedge}b \times' a^2b^2 \tag{5.1}$$

This theorem may readily be tri-alized.

Since $N = C$ for Boolean rings (= 2 rings), if one compares (5.1) with the corresponding formula for 2-rings, namely

$$a + b = ab^{\wedge} \times' a^{\wedge}b$$

it is found that neither "covers" the other, that is, neither reduces to the correct formula for $+$ in the other ring. It is conjectured that no (other) equational and logical definitions of $+$ for $p = 2$ and $p = 3$ exist which cover each other, and similar conjectures seem reasonable for $p$-rings and $p'$-rings, for $p \neq p'$.

Exactly as the logic of propositions (= 2-valued logic) is mathematically equivalent to the simplest 2-ring (= field-logic $F_2$), so the 3-valued logic is mathematically equivalent to the simplest 3-ring, the field-logic $F_3$. In 3-valued logic the "logical language" is $(F_3, \times, \times', \times'', {}^{\wedge}, {}^{\vee})$, while the "ring language" is $(F_3, +, \times)$. The tri-ality theorem reigns over both of these, as it does over any 3-ring-logic, and throws new light on the 3-valued logic. We cannot enter into the details of this here.

One of a number of basic questions opened up by the foregoing theory and not yet completely settled, is the following: Given a ring $R$, is there at least one group $K$ which is *fully adapted* to $R$, that is, which will convert $R$ into a ring-logic $(K)$?

[1] Foster, A. L., "The Theory of Boolean-Like Rings," *Trans. Am. Math. Soc.*, **59**, 166–187 (1946).

[2] Foster, A. L., "The Idempotent Elements of a Commutative Ring Form a Boolean Algebra; Ring Duality and Transformation Theory," *Duke Math. J.*, **12**, 143–152 (1945).

[3] "Maximal Idempotent Sets in a Ring with Unit," *Ibid.*, **13**, 247–258 (1946).

[4] Foster, A. L., "On the Permutational Representation of General Sets of Operations by Partition Lattices," to appear in *Trans. Am. Math. Soc.*, June, 1949.

[5] Foster, A. L., and Bernstein, B. A., "Symmetric Approach to Commutative Rings with Duality Theorem: Boolean Duality as Special Case," *Duke Math. J.*, **11**, 603–616 (1944).

[6] Foster, A. L., and Bernstein, B. A., "A Dual Symmetric Definition of Field," *Am. J. Math.*, **67**, 329–349 (1945).

[7] Harary, F., "The Structure of Boolean-Like Rings," in process of publication.

---

# *ON A THEOREM OF DIEUDONNÉ*

BY PAUL R. HALMOS

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CHICAGO

Communicated by A. A. Albert, November 12, 1948

1. The fact that a measure preserving transformation may be decomposed into ergodic parts was proved first by von Neumann, under the assumption that the domain of the transformation is an $m$-space (= a complete, separable, metric space with a regular measure).[1] Several years later I defined the concept of direct sum for measure spaces and, in terms of this concept, proved a more general decomposition theorem for measure spaces subject to certain separability restrictions.[2] My proof rested on a lemma, stated by Doob, concerning the existence of certain measures on such measure spaces. It was recently pointed out by Dieudonné that Doob's lemma and my decomposition theorem are both false.[3] The situation, in greater detail, is that my proof of the decomposition theorem is valid for a measure space if and only if the conclusion of Doob's lemma is correct for that space. An examination of Doob's proof shows that his conclusion is certainly correct if the measure space in question is the unit interval. Since, however, there is a wide class of measure spaces (called normal) for which it is known that they may be put into one to one measure preserving correspondence with the unit interval,[4] a large part of my original decomposition theorem is saved—all that has to be done is to replace the original separability assumption by the assumption of normality.

Since every non-atomic $m$-space is normal, and since the few additional considerations required by the possibly atomic case are trivial, my treatment for normal spaces still constitutes a generalization of von Neumann's theorem and a simplification of his proof.

In addition to finding the error in the decomposition theorem as originally stated, Dieudonné has also proposed two partial replacements of it.[5] Since the second of these deals with compact (and *a fortiori* complete) metric spaces, it follows from the remarks in the preceding paragraph that that theorem is in fact a special case of the decomposition theorem for normal spaces. Dieudonné's first theorem, however, is a very beautiful and powerful addition to decomposition theory. Since his proof of this theorem rests on some earlier theorems of his concerning Banach space valued integrals, I propose in this note to offer another, somewhat simpler, and more direct proof and to make some comments about the relation of the result to previously known facts.

2. Let $Y$ be a totally disconnected, compact Hausdorff space and let $\mathbf{E}$ be the class of all open-closed subsets of $Y$. Let $\mathbf{B}$ be the smallest Boolean $\sigma$-algebra of sets containing $\mathbf{E}$; the sets in $\mathbf{B}$ are called the *Baire sets* of $Y$. A *Baire measure* in $Y$ is a finite (countably additive) measure on $\mathbf{B}$. For every Baire measure $\mu$ in $Y$ and for every Baire set $B$,

$$\mu(B) = \sup\{\mu(K): B \supset K \in \mathbf{K}\} = \inf\{\mu(U): B \subset U \in \mathbf{U}\},$$

where $\mathbf{K}$ is the class of all compact Baire sets and $\mathbf{U}$ is the class of all open Baire sets; these relations are usually expressed by saying that every Baire measure is regular.[6] If an open-closed subset of $Y$ is the union of a class of open-closed sets, then, by compactness, all but a finite number of members of the class must be empty. It follows that every non-negative, finite and finitely additive set function defined on the class of all open-closed subsets of $Y$ is in a vacuous sense countably additive on this class and hence that it has a unique extension to a Baire measure in $Y$. A Baire measure $\mu$ in $Y$ is *positive* if $\mu(E) > 0$ for every $E$ in $\mathbf{E}$ (and hence $\mu(U) > 0$ for every $U$ in $\mathbf{U}$). If the space $Y$ has the property that the class $\mathbf{E}$ of all open-closed sets is a $\sigma$-lattice, and if $\mu$ is a positive Baire measure in $Y$, then the pair $(Y, \mu)$ will be called a *Kakutani space*. Equivalently, a Kakutani space is a totally disconnected, compact Hausdorff space with a positive Baire measure and with the property that corresponding to any countable class of open-closed sets there is a least open-closed set containing them all. If $B$ is a Baire set in a Kakutani space $(Y, \mu)$, then there exists a unique open-closed set $E$ in $Y$ such that $\mu(B \Delta E) = 0$ (where $B \Delta E = (B - E) \cup (E - B)$ is the symmetric difference of $B$ and $E$). A Boolean $\sigma$-algebra $\mathbf{J}$ of Baire sets (i.e., a $\sigma$-subalgebra of the algebra $\mathbf{B}$ of all Baire sets) will be called *full* if for each $B$ in $\mathbf{J}$ the open-closed set $E$ for which $\mu(B \Delta E) = 0$ is also contained in $\mathbf{J}$.

Let $\mathbf{A}$ be an abstract Boolean $\sigma$-algebra, let $\lambda$ be a finite measure on $\mathbf{A}$, and denote by $\mathbf{A}/\lambda$ the quotient algebra of $\mathbf{A}$ modulo the $\sigma$-ideal of all those elements $A$ for which $\lambda(A) = 0$. By Stone's theorem there exist a totally disconnected compact Hausdorff space $Y$ and a one-to-one mapping $\Phi$ (the Stone mapping) from $\mathbf{A}/\lambda$ onto the class $\mathbf{E}$ of all open-closed subsets of $Y$ such that $\Phi$ preserves all finite Boolean operations. If $\mu(E)$ is defined for every $E$ in $\mathbf{E}$, by $\mu(E) = \lambda(\Phi^{-1}(E))$ then $\mu$ has a unique extension to a positive Baire measure in $Y$. It is easy to verify that $(Y, \mu)$ is a Kakutani space; it will be called the Kakutani space *associated* with $\mathbf{A}$ and $\lambda$. It may be remarked that by combining $\Phi$ with the natural homomorphism from the algebra $\mathbf{A}$ onto the quotient algebra $\mathbf{A}/\lambda$ one obtains a homomorphic mapping of $\mathbf{A}$ onto $\mathbf{E}$; it is convenient and harmless to use the same symbol, i.e., $\Phi$, for this mapping as for the original isomorphism. While with the new interpretation of $\Phi$ it is no longer true that if $\Phi(A_1) = \Phi(A_2)$, then $A_1 = A_2$, a similar and equally useful statement is still true: if $\Phi(A_1) = \Phi(A_2)$, then $A_1 = A_2\,[\lambda]$, where the equality of $A_1$ and $A_2$ modulo $\lambda$ is of course defined by $\lambda(A_1 \Delta A_2) = 0$.

The construction in the preceding paragraph is of particular interest when applied to a finite measure space $(X, \mathbf{A}, \lambda)$, i.e., when $\mathbf{A}$ is a Boolean $\sigma$-algebra of subsets of a set $X$. In this case it may be shown[7] that the mapping $\Phi$, when viewed as a mapping between characteristic functions, has a unique extension, which may also be denoted by $\Phi$ and which maps each essentially bounded, real valued, measurable function $f$ on $X$ into a real valued, continuous function $g$ on $Y$ so that

(i) $\Phi(\alpha_1 f_1 + \alpha_2 f_2) = \alpha_1\Phi(f_1) + \alpha_2\Phi(f_2)$,

(ii) $\Phi(f_1 f_2) = \Phi(f_1)\Phi(f_2)$,

(iii) if $f \geqq 0\ [\lambda]$, then $\Phi(f) \geqq 0$, and

(iv) if $\Phi(f) = g$, then $\|f\|_\infty = \|g\|$, where

$\|f\|_\infty = \inf\{k: \lambda(\{x: |f(x)| > k\}) = 0\}$ and $\|g\| = \inf\{k: \{y: |g(y)| > k\} = 0\}$.

3. In the language and notation introduced above, Dieudonné's fundamental decomposition theorem may be stated as follows.

THEOREM. *If $(Y, \mu)$ is a Kakutani space and if $\mathbf{J}$ is a full Boolean $\sigma$-algebra of Baire sets, then there exist a Kakutani space $(Z, \nu)$, a continuous mapping $\pi$ from $Y$ onto $Z$, and, for each $z$ in $Z$, a Baire measure $\mu^z$ in $\pi^{-1}(z)$ such that the set transformation induced by $\pi$ is a one-to-one mapping from $\mathbf{J} \cap \mathbf{E}$ onto the class $\mathbf{F}$ of all open-closed sets in $Z$ and such that, for every open-closed set $E$ in $Y$ (i.e., for every $E$ in $\mathbf{E}$),*

$$\mu(E) = \int \mu^z(E \cap \pi^{-1}(z))d\nu(z).^8$$

*Proof:* Let $(Z, \nu)$ be the Kakutani space associated with $\mathbf{J}$ and $\mu$ and denote the corresponding Stone mapping (from sets and functions of $Y$ onto sets and functions of $Z$) by $\Pi$. Consider for each point $y$ in $Y$ the class of all those sets in $\mathbf{J} \cap \mathbf{E}$ which contain $y$. Since $\Pi$ is a Boolean isomorphism between $\mathbf{J} \cap \mathbf{E}$ and $\mathbf{F}$, the image under $\Pi$ of this class of sets has the finite intersection property and, consequently, the sets of this image class have a non-empty intersection. Since, moreover, any two distinct points of $Z$ may be separated by disjoint open-closed sets, it follows that this intersection consists of exactly one point $z$, and that therefore a transformation $\pi$ from $Y$ into $Z$ is uniquely defined by $\pi(y) = z$. It is easy to see that $\pi$ maps $Y$ onto $Z$ and, since the set transformation induced by $\pi$ coincides with $\Pi$, that $\pi$ is continuous. All that remains to be done is to construct the measures $\mu^z$ and to establish the integral formula.

Consider, for each $E$ in $\mathbf{E}$, the set function $\mu_E$ on $\mathbf{J}$ defined, for $J$ in $\mathbf{J}$, by $\mu_E(J) = \mu(E \cap J)$. Since $\mu_E$ is absolutely continuous with respect to $\mu$, it follows from the Radon-Nikodym theorem that there exists a bounded, non-negative function $p_E$ on $Y$ such that $p_E$ is measurable with respect to $\mathbf{J}$ and such that

$$\mu(E \cap J) = \int_J p_E(y)d\mu(y)$$

for every $J$ in $\mathbf{J}$. Write $q_E = \Pi(p_E)$; it is sometimes convenient to write $q(E, z)$ or $q^z(E)$ for $q_E(z)$. It is clear from the definition of $q$ that

$$\mu(E \cap J) = \int_{\Pi(J)} q(E, z)d\nu(z)$$

for every $E$ in $\mathbf{E}$ and $J$ in $\mathbf{J}$.

Since $p_{E_1}(y) + p_{E_2}(y) = p_{E_1 \cup E_2}(y)$ $[\mu]$ whenever $E_1$ and $E_2$ are disjoint open-closed subsets of $Y$, it follows from the fact that $q$ is a continuous function of its second argument that the set function $q^z$ is finitely additive on $\mathbf{E}$ for each $z$ in $Z$. Each set function $q^z$ has therefore a unique extension, which may also be denoted by $q^z$, to a Baire measure in $Y$.

The proof is almost complete. If $\mu^z$ is defined, for each Baire set $B_z$ in $\pi^{-1}(z)$, by $\mu^z(B_z) = q(B_z, z)$, then it is clear that $\mu^z$ is a Baire measure. If it were known that, for every $E$ in $\mathbf{E}$,

$$\mu^z(E \cap \pi^{-1}(z)) = q(E, z),$$

the desired integral formula would follow from the defining properties of the functions $p$ and $q$. Since the last written relation is equivalent to

$$q(E, z) = q^z(E \cap \pi^{-1}(z)),$$

the proof can be completed by showing that $q^z(E - \pi^{-1}(z)) = 0$ for every $z$ in $Z$. Since $q^z$ is a Baire measure, and therefore regular, it is sufficient to prove that $q^z(K) = 0$ for every compact Baire set $K$ which is disjoint from $\pi^{-1}(z)$. Since $\pi(K)$ is a compact subset of $Z$ which does not contain $z$, there exist two disjoint open-closed sets $F$ and $G$ in $Z$ such that $\pi(K) \subset F$

and $z \in G$. Since $\pi^{-1}(F) \in \boldsymbol{E}$ and $\pi^{-1}(G) \in \boldsymbol{J}$ (in fact both sets $\pi^{-1}(F)$ and $\pi^{-1}(G)$ belong to both classes $\boldsymbol{E}$ and $\boldsymbol{J}$), it follows that

$$0 = \mu(\pi^{-1}(F) \cap \pi^{-1}(G)) = \int_G q(\pi^{-1}(F), z)d\nu(z).$$

Since $q$ is a continuous function of its second argument, it follows, finally, that $q^z(K) \leqq q^z(\pi^{-1}(F)) = q(\pi^{-1}(F), z) = 0$, and the proof of the theorem is complete.

4. One of the most important applications of this result is to the case in which $\boldsymbol{J}$ is the class of all Baire sets invariant under a measure preserving homeomorphism $S$ of $Y$ onto itself; in this case the induced transformations $S_z$ on the spaces $\pi^{-1}(z)$ are the ergodic components of $S$. If $T$ is a one to one measure preserving transformation of a finite measure space $(X, \boldsymbol{A}, \lambda)$ onto itself, then there exists a measure preserving homeomorphism $S$ on the associated Kakutani space $Y$ such that $A_1 = TA_2$ $[\lambda]$ if and only if $\Phi(A_1) = S\Phi(A_2)$.[9] In other words, every measure preserving transformation may be represented, modulo sets of measure zero, as a measure preserving homeomorphism on a Kakutani space, and consequently Dieudonné's theorem yields a solution of the general problem of decomposing a measure preserving transformation into ergodic parts. This solution is better than previous solutions in that it does not depend on any assumptions of separability and in that it applies not only to cyclic and one-parameter groups but to arbitrary groups of measure preserving transformations.

[1] *Ann. Math.*, **33**, 587–642 (1932).

[2] *Duke Math. J.*, **8**, 386–392 (1941).

[3] *Ann. Univ. Grenoble*, **23**, 25–53 (1948). In the sequel this paper will be denoted by D.

[4] *Ann. Math.*, **43**, 332–350 (1942).

[5] D, Theorems 4 and 5.

[6] A proof of this statement and a complete discussion of the concept of regularity will appear in my forthcoming book entitled *Measure Theory*.

[7] D, p. 27.

[8] In Dieudonné's formulation of this theorem an arbitrary real valued, continuous function appears in place of the arbitrary open-closed set $E$. It is easy to reconstruct his statement from mine by means of a familiar approximation argument.

[9] D, p. 51.

# *OSCULATING CONICS OF THE INTEGRAL CURVES OF THIRD ORDER DIFFERENTIAL EQUATIONS OF THE TYPE (G)*[*]

By Edward Kasner and John De Cicco

Departments of Mathematics, Columbia University, New York, and Illinois Institute of Technology, Chicago

Communicated November 1, 1948

1. *Third-Order Differential Equations of the Type* (*G*).—The importance in geometry and physics of third order differential equations of the type (*G*),

$$y''' = G(x, y, y')y'' + H(x, y, y')y''^2 \qquad (1)$$

has been brought out in previous papers.[1]

Perhaps the most important case is the system $S_k$ of $\infty^3$ curves in a positional or generalized field of force.[2] A system $S_k$ consists of curves along which a constrained motion is possible so that the pressure $P$ is proportional to the normal component $N$ of the force vector. Thus $P = kN$ where $k(\neq -1)$ is the constant factor of proportionality.

The systems $S_k$ of physical interest are (*a*) dynamical trajectories where $k = 0$, (*b*) general catenaries where $k = 1$, (*c*) brachistochrones where $k = -2$, and (*d*) velocity curves where $k = \infty$.

Other cases of the type (*G*) are the $\infty^3$ plane sections of a surface in ordinary space[3] and the $\infty^3$ curvature trajectories of a given family of $\infty^2$ curves in the plane.[4]

2. *Summary of Results.*—The systematic development of the osculating parabolas (four point contact) of systems of $\infty^3$ curves of the type (*G*) with special emphasis on the systems $S_k$ has been given by Kasner. In the present paper, we shall begin the study of the osculating conics (five point contact) for the type (*G*).

The $\infty^1$ osculating conics constructed at a lineal element $E$ to the $\infty^1$ curves of a system of the type (*G*) passing through $E$, not only pass through $E$, but also touch another conic section in general position. Their centers describe still another conic section passing through the point of $E$. The foci of these conic sections describe an algebraic curve of the sixth degree with a singular point of fourth order at the point of $E$.

3. *The Osculating Conics of a System of* $\infty^3$ *Curves of the Type* (*G*).—The osculating conics of a single curve $y = f(x)$ are given by

$$9y''^4X^2 + 6y''^2y'''X(Y - y'X) + (3y''y^{\text{iv}} - 4y'''^2)(Y - y'X)^2 - 18y''^3(Y - y'X) = 0, \qquad (2)$$

where $(X, Y)$ are the running coördinates of a point on the conic relative to the point $(x, y = f(x))$ on the curve.

Upon substituting (1) into (2), it is found that *the osculating conics of a system of* $\infty^3$ *curves of the type* (*G*) *are*

$$9y''^2X^2 + 6y''(G + Hy'')X(Y - y'X) + (\alpha + \beta y'' + \gamma y''^2) \times \\ (Y - y'X)^2 - 18y''(Y - y'X) = 0, \quad (3)$$

where

$$\begin{aligned} \alpha &= -G^2 + 3(G_x + y'G_y), \\ \beta &= GH + 3(G_{y'} + H_x + y'H_y), \\ \gamma &= 2H^2 + 3H_{y'}. \end{aligned} \quad (4)$$

The conic (3) is degenerate if and only if the curvature of the curve is zero. In this case, the conic becomes the line $Y = y'X$ counted twice.

The conic (3) is an ellipse, parabola, hyperbola according as the expression,

$$(G + Hy'')^2 - (\alpha + \beta y'' + \gamma y''^2), \quad (5)$$

is negative, zero, positive. Thus at a lineal element $E(x, y, y')$, there are at most two integral curves which are hyperosculated by their osculating parabolas.

In the class of all differential equations of third order of the type ($G$), the subclass for which $\gamma = H^2$, or $H^2 + 3H_y = 0$, is of particular interest as this subclass contains the systems of $\infty^3$ dynamical trajectories of a positional field of force.

In this particular subclass, there is, at most, one integral curve through a lineal element $E$ which is hyperosculated by its osculating parabola.

4. *The Envelope of the Family of Conics* (3).—The equation (3) is quadratic in $y''$. Upon setting the discriminant of the quadratic equation in $y''$ equal to zero, the following result is obtained.

THEOREM 1. *Consider the* $\infty^1$ *integral curves of a third order differential equation of the type* (*G*) *which pass through a lineal element E. The* $\infty^1$ *osculating conics constructed at E to these curves, not only pass through E, but also touch the conic section*

$$36(\alpha - G^2)X^2 + 12(2H\alpha - G\beta)X(Y - y'X) + (4\alpha\gamma - \beta^2)(Y - y'X)^2 + \\ 36[6GX + \beta(Y - y'X)] - 324 = 0. \quad (6)$$

The conic section (6) is degenerate only in the following two cases:

(*a*) *The case where* $\alpha = 0$, *or* $3(G_x + y'G_y) = G^2$. In this case, the conics (3) form a linear family in point coördinates. They are all tangent at the lineal element $E$ and pass through two other points given as the intersections of the following loci

$$9X^2 + 6HX(Y - y'X) + \gamma(Y - y'X)^2 = 0, \\ 6GX + \beta(Y - y'X) - 18 = 0. \quad (7)$$

(*b*) *The case where* $\gamma = H^2$, *or* $H^2 + 3H_{y'} = 0$. Here the conics (4) form a linear family in line coördinates. They are all tangent at the lineal element $E$ and are tangent to the two straight lines

$$\pm 2\alpha^{1/2}[3X + H(Y - y'X)] = 6GX + \beta(Y - y'X) - 18. \quad (8)$$

These two straight lines intersect in the point

$$X + i(Y - y'X) = 6(-H + 3i)/(\beta - 2GH). \quad (9)$$

The two lines (8) may be real and distinct, or real and coincident, or conjugate imaginary, or the line at infinity.

5. *The Central Locus of the Family of Conics* (*3*).—The equations for the center $(X, Y)$ of the conic (3) are

$$3y''[3y'' - y'(G + Hy'')]X + [3y''(G + Hy'') - y'(\alpha + \beta y'' + \gamma y''^2)] \times (Y - y'X) + 9y'y'' = 0, \quad (10)$$
$$3y''(G + Hy'')X + (\alpha + \beta y'' + \gamma y''^2)(Y - y'X) - 9y'' = 0.$$

THEOREM 2. *The centers of the* $\infty^1$ *conics described in Theorem 1, generate the conic section*

$$9(\alpha - G^2)X^2 + 3(2\alpha H - \beta G)X(Y - y'X) + (\gamma G^2 + \alpha H^2 - \beta GH) \times (Y - y'X)^2 + 9G[3X + H(Y - y'X)] = 0, \quad (11)$$

*which passes through the point of E in the direction of slope* $y' - 3/H$.

The conic section (11) is degenerate only in the following two cases:

(*a*) *The case where* $G = 0$. Here the conic section (11) is the straight line

$$3X + H(Y - y'X) = 0, \quad (12)$$

counted twice.

(*b*) *The case where* $\gamma = H^2$, *or* $H^2 + 3H_{y'} = 0$. In this case, the conic section (11) consists of the straight line (12) and the straight line

$$3(\alpha - G^2)X + [H(\alpha + G^2) - \beta G](Y - y'X) + 9G = 0. \quad (13)$$

6. *The Focal Locus of the Family of Conics* (*3*). There is no loss in generality by taking $y' = 0$ in (3). The foci of the conic (3) is given by the equation in $Z = X + iY$,

$$[(\alpha + \beta y'' + \gamma y''^2) - (G + Hy'')^2]Z^2 + 6[G + Hy'' - 3iy''] \times Z - 9 = 0 \quad (14)$$

THEOREM 3. *The foci of the* $\infty^1$ *conics described in Theorem 1, generate the algebraic curve of sixth degree*

$$(\alpha - G^2)(X^2 + Y^2)^2(3X + HY)^2 - (\beta - 2GH)Y(3X + HY) \times (X^2 + Y^2)[G(X^2 + Y^2) - 3X] + (\gamma - H^2)Y^2[G(X^2 + Y^2) - 3X]^2 + 9(3X + HY)(X^2 + Y^2)[2G(X^2 + Y^2) - 3X + HY] = 0, \quad (15)$$

*with double points at the circular points I and J at infinity and a singular point of fourth order at the point of E.*

In the case where $\gamma = H^2$ or $H^2 + 3H_{y'} = 0$, it is found that the focal locus is the cubic curve

$$(\alpha - G^2)(X^2 + Y^2)(3X + HY) - (\beta - 2GH)Y[G(X^2 + Y^2) - 3X] + 9[2G(X^2 + Y^2) - 3X + HY] = 0, \quad (16)$$

with two minimal tangent lines and an ordinary point at the point of $E$.

The cubic curve (16) is degenerate only when

$$\beta - 2GH = \frac{2H}{G}(\alpha - G^2) = -\frac{2G}{H}(H^2 + 9). \quad (17)$$

In this case it consists of the straight line $3X - HY = 0$, and the null circle with center at

$$X + iY = \frac{3H(H - 3i)}{G(H^2 + 9)}. \quad (18)$$

The simplest physical theorem of the present paper is that *the locus of centers of the osculating conics connected with dynamical trajectories* is always *a straight line; and these osculating conics touch two straight lines which intersect on the line of force.* This is easily verified for the Newtonian field (Kepler orbits).

---

[*] Presented to the American Mathematical Society, 1948.

[1] Kasner, "Differential-Geometric Aspects of Dynamics," *Amer. Math. Soc. Publications*, 1913, 1934. See a series of papers in the *Trans. Amer. Math. Soc.*, 1906–1910.

[2] Kasner and De Cicco, "Physical Curves in Generalized Fields of Force," PROC. NAT. ACAD. SCI., **33**, 169–172 (1948).

[3] Kasner, "Dynamical Trajectories and the $\infty^3$ Plane Sections of a Surface," *Ibid.*, **17**, 370–376 (1931).

[4] Kasner, "Dynamical Trajectories and Curvature Trajectories," *Bull. Amer. Math. Soc.*, **44**, 449–455 (1934). De Cicco, "Dynamical and Curvature Trajectories in Space," *Trans. Amer. Math. Soc.*, **57**, 270–286 (1945).

---

# *REMARK CONCERNING INTEGRATION*

BY E. J. MCSHANE

SCHOOL OF MATHEMATICS, UNIVERSITY OF VIRGINIA

Communicated November 23, 1948

M. H. Stone has communicated[1] a scheme of defining an integral, closely related to the Daniell definition[2] and affording unusually simple proofs of the main theorems. However, with this definition it seems impossible to prove in full strength the theorem on the representation in integral form of the general bounded linear functional over the space of continuous functions on a locally compact Hausdorff space. The purpose of the present note is to propose a modification of Stone's definition which introduces no new difficulties and removes the difficulty just mentioned.

Our postulates are the following, closely resembling those of Daniell and those of Stone.

(1$a$) *$\mathfrak{E}$ is a non-void class of real functions on a non-void domain $X$, such that whenever $r$ is real and $\varphi$ and $\psi$ are in $\mathfrak{E}$, the functions $r\varphi$, $\varphi + \psi$ and* sup $(\varphi, \psi)$ *are also in* $\mathfrak{E}$. (The last is the function whose value at $x \in X$ is the greater of $\varphi(x)$ and $\psi(x)$.) It follows that if $\varphi \in \mathfrak{E}$, so is $|\varphi| = \sup(\varphi, -\varphi)$.

(1$b$) *$E$ is a real function on $\mathfrak{E}$ such that if $r$ is real and $\varphi$ and $\psi$ are in $\mathfrak{E}$, $E(r\varphi) = rE(\varphi)$, $E(\varphi + \psi) = E(\varphi) + E(\psi)$, $E(|\varphi|) \geqq 0$.*

If as usual we understand $\varphi \geqq \psi$ to mean $\varphi(x) \geqq \psi(x)$ for all $x \in X$, the class of all real functions on $X$ is partially ordered. More than this, it is directed, for if $f$, $g$ are real on $X$, then $\sup(f, g) \geqq f$ and $\geqq g$. Likewise, by (1$a$) the set $\mathfrak{E}$ is directed. If $\Delta$ is a directed set of functions $f$ on $X$, we denote by $\lim(f|\Delta, \geqq)$ the function whose value at $x \in X$ is the Moore-Smith[3] limit of the functional values $f(x)$ $(f \in \Delta)$ with respect to the partial ordering $\geqq$. Likewise, if $\Delta \subset \mathfrak{E}$ we define $\lim(E(f)|\Delta, \geqq)$ as the corresponding Moore-Smith limit. These limits coincide with $\sup(\varphi|\varphi \in \Delta)$ and sup $(E(\varphi)|,\varphi \in \Delta)$, respectively.

Our second postulate is

(2) *If $\psi \in \mathfrak{E}$ and $\Delta$ is a directed subset of $\mathfrak{E}$ such that* $\lim(\varphi|\Delta, \geqq) \geqq \psi$, *then* $\lim(E(\varphi)|\Delta, \geqq) \geqq E(\psi)$. This is easily seen to be equivalent to

(3) *If $\psi \in \mathfrak{E}$ and $\Delta$ is a directed set of non-negative $\varphi \in \mathfrak{E}$, and* $\lim(\varphi|\Delta, \geqq) \geqq \psi$, *then* $\lim(E(\varphi)|\Delta, \geqq) \geqq E(\psi)$. For if (3) holds and $\Delta$ is as in (2), let $\varphi_0 \in \Delta$, and let $\Delta_0$ be the set $[\varphi - \varphi_0|\varphi \in \Delta, \varphi \geqq \varphi_0]$. This is a directed subset of $\mathfrak{E}$ with limit $\geqq \psi - \varphi_0$. By (3), $\lim(E(\varphi - \varphi_0)|\varphi \in \Delta_0) \geqq E(\psi - \varphi_0)$, and (2) holds.

If $f$ is extended-real (allowing $+\infty$ and $-\infty$ as functional values), and there is a directed set $\Delta$ of non-negative $\varphi \in \mathfrak{E}$ such that $\lim(\varphi|\Delta, \geqq) \geqq |f|$, we define the norm $N(f)$ to be

(4) $N(f) = \inf \lim(E(\varphi)|\Delta, \geqq)|\Delta$ *a directed* set of non-negative $\varphi \in \mathfrak{E}$ *such that* $\lim(\varphi|\Delta, \geqq) \geqq |f|)$;

if no such directed set $\Delta$ exists, we define $N(f) = +\infty$.

The first, second, fourth and fifth of the properties

(5) $0 \leqq N(f) \leqq \infty$,

(6) *$N(rf) = |r|N(f)$ unless $r = 0$ and $N(f) = +\infty$,*

(7) $|f| \leqq \sum_1^{\infty} |f_n|$ *implies* $N(f) \leqq \sum_1^{\infty} N(f_n)$,

(8) $N(|f|) = N(f)$,

(9) $\varphi \in \mathfrak{E}$ *implies* $N(\varphi) = E(|\varphi|)$

are easily established; the last requires the use of (2). To establish (7), let $\epsilon$ be an arbitrary positive number, and for each positive integer $n$ let $\Delta_n$ be a directed subset of $\mathfrak{E}$ such that $\lim(\varphi|\Delta_n, \geqq) \geqq |f_n|$ and $\lim(E(\varphi)|\Delta_n, \geqq) \geqq N(f_n) + 2^{-n}\epsilon$. Let $\Delta$ be the aggregate of sums of finitely many functions all chosen from different $\Delta_n$. Since each $\Delta_n$ is directed, so is $\Delta$. It is easy to show that $\lim(\varphi|\Delta, \geqq) \geqq |f|$ and $\lim(E(\varphi)|\Delta, \geqq) \leqq \sum_{n=1}^{\infty} N(f_n) + \epsilon$, whence (7) follows.

From this point on, Stone's paper can be followed without change. $\mathfrak{F}$ is defined to be the class of functions on $X$ with finite norm, functions $f$ and $g$ being identified if $N(f - g) = 0$. $\mathfrak{L}$ is the closure of $\mathfrak{E}$ in $\mathfrak{F}$, and the definition of the integral is equivalent to defining $L(f)$ $(f \in \mathfrak{L})$ as $\lim E(\varphi)$ $(\varphi \in E;$ $N(f - \varphi) \to 0)$.

We now establish a lemma exhibiting a class of summable functions.

(10) *Let* $\Delta$ *be a subset of* $\mathfrak{E}$, *directed with respect to* $\geqq$, *on which* $E$ *is bounded. Define* $f = \lim(\varphi|\Delta, \geqq)$. *Then* $f$ *is summable, and its integral is* $L(f) = \lim(E(\varphi)|\Delta, \geqq)$. *Herein the order relation* $\geqq$ *may be everywhere replaced by* $\leqq$.

For let $\epsilon$ be positive. There is a $\psi$ in $\Delta$ such that $E(\psi) > \lim(E(\varphi)|\Delta, \geqq) - \epsilon$. Let $\Delta_0$ be the aggregate of functions $\delta = \varphi - \psi$, where $\varphi \in \Delta$ and $\varphi \geqq \psi$. This is a directed subset of $\mathfrak{E}$, and $\lim(\delta|\Delta_0, \geqq) = f - \psi$, and $\lim(E(\delta)|\Delta_0, \geqq) = \lim(E(\varphi)|\Delta, \geqq) - E(\psi) < \epsilon$. Hence $N(f - \psi) < \epsilon$, and $f$ is in the closure $\mathfrak{L}$ of $\mathfrak{E}$, that is, $f$ is summable. The statement concerning $\leqq$ follows by change of sign.

One might suspect that (10) could be generalized by letting $\Delta$ be a directed subset of $\mathfrak{L}$ on which $L$ is bounded. This is false. Let $X$ be the interval $[0, 1]$, $\mathfrak{E}$ the class of continuous functions, $E(\varphi)$ the Riemann integral of $\varphi$; then $L(.)$ is the Lebesgue integral. Let $f$ be a positive-valued non-measurable function, $\Delta$ the class of functions $\leqq f$ on $X$ which vanish except on a finite subset of $X$. Then $\Delta$ is a directed subset of $\mathfrak{L}$ and $L$ vanishes on $\Delta$; but $\lim(\varphi|\Delta, \geqq) = f$, which is not summable.

If all directed sets $\Delta$ in the preceding definitions are required to be countable, our definition reduces to that of Stone; if $\Delta$ is directed with respect to $\geqq$ and is countable, $\Delta = [\varphi_1, \varphi_2, \ldots]$, we define $\psi_n = \sup(\varphi_1, \ldots, \varphi_n)$, and then the sequential limits of $\psi_n$ and $E(\psi_n)$ are the same as $\lim(\varphi|\Delta, \geqq)$ and $\lim(E(\varphi)|\Delta, \geqq)$, respectively. In many examples, there is a countable subset $\mathfrak{E}_0$ of $\mathfrak{E}$ such that for each $\varphi \in \mathfrak{E}$ there is a directed subset $\Delta_0$ of $\mathfrak{E}_0$ such that $\lim(\psi | \psi \epsilon \Delta_0) = \varphi$ and $\lim(E(\psi) | \psi \epsilon \Delta_0) = E(\varphi)$. Under these circumstances each directed subset $\Delta$ of $\mathfrak{E}$ can be replaced by a countable subset having the same limits for $\varphi$ and $E(\varphi)$. Then the present definition coincides with Stone's.

If $X$ is $[0, 1]$, $\mathfrak{E}$ the continuous functions on $X$ and $E$ the Riemann integral, our definition and Stone's yield $L(.)$ as the Lebesgue integral. But if $\mathfrak{E}$ consists of all step-functions, the preceding example shows that our (2) is not satisfied. Stone's (2) holds, and he again obtains the Lebesgue integral.

Now let $X$ be a locally compact Hausdorff space and $\mathfrak{E}$ the collection of functions continuous on $X$ and vanishing except on a compact set. Then $\mathfrak{E}$ satisfies (1*a*). Let $E$ on $\mathfrak{E}$ satisfy (1*b*). To show that (3) holds, denote by $K$ the closure of the set of $x$ on which $\psi(x) > 0$. For each positive $\epsilon$, there is a non-negative $\varphi_0 \epsilon \mathfrak{E}$ which is positive on $K$ and has $E(\varphi_1) < \epsilon$. Since $K$ is compact, there are finitely many $\varphi_i \epsilon \Delta$ such that the sets $[x | \varphi_i(x) > \psi(x) - \varphi_0(x)]$ cover $K$, whence $\lim(E(\varphi) | \Delta, \geqq) \geqq E(\psi) - \epsilon$, and (3) follows. Now the preceding discussion guarantees the existence of a functional $L(.)$ with the fundamental properties of the Lebesgue integral.

In his second note Stone has defined a concept of measurability of functions such that $f \geqq 0$ is measurable if and only if $\min(f, g)$ is summable for all summable $g \geqq 0$. The measurable sets are those with measurable characteristic function; the measure is the integral of the characteristic function when this exists. This measure possesses the properties commonly expected. We wish to show that *every Borel set is measurable*. It suffices to show that every open set $G$ has a measurable characteristic function $f_G$. For $x_0$ in $G$, there is a function $\varphi \epsilon \mathfrak{E}$ such that $0 \leqq \varphi \leqq f_G$, $\varphi(x_0) = 1$. So the set $\Delta$ of all non-negative elementary functions $\leqq f_G$ is a directed system with limit $f_G$. Let $g$ be non-negative and summable, $n$ a positive integer; then there is a directed system $\Delta_n$ of non-negative functions of $\mathfrak{E}$ such that $g_n = \lim(\varphi | \Delta_n, \geqq) \geqq g$ and $\lim(E(\varphi) | \Delta_n, \geqq) < L(g) + 1/n$. The set $\Delta_n^*$ of all functions $\min(h, k)$ $h \epsilon \Delta_n$, $k \epsilon \Delta$ is directed with respect to $\geqq$ and its limit is $\min(f_G g_n)$, which is summable by (10). By the majorized convergence theorem $\min(f_G g_0)$ is summable, where $g_0 = \lim_{n \to \infty} \min(g_1, \ldots, g_n)$. But $g_0 \geqq g$ and $L(g_0) = L(g)$, so $g_0$ and $g$ are almost everywhere equal. Hence $\min(f_G, g)$ is summable, and $G$ is measurable.

If $K$ is compact there is a non-negative function $\varphi$ of $\mathfrak{E}$ which is $\geqq 1$ on $K$; $K$ is measurable, being a Borel set, so it has finite measure.

---

[1] Stone, M. H., PROC. NAT. ACAD. SCI., **34**, 336–342, 447–455, 483–490 (1948).

[2] Daniell, P. J., *Ann. Math.*, **19**, 279–294 (1917).

[3] Moore, E. H., and Smith, H. L., *Amer. Jour. Math.*, **44**, 102–121 (1922).

# NOTES ON INTEGRATION, IV

By M. H. Stone

Department of Mathematics, The University of Chicago

Communicated September 4, 1948

The theory of integration presented in the earlier notes[1] postulates an elementary integral in terms of which all other concepts are defined. We must therefore undertake to compare the general integrals and other mathematical objects which are associated with different elementary integrals. In this note we offer such a comparison, giving the appropriate general forms of such classical results as the Lebesgue decomposition theorem and the Radon-Nikodym theorem. This is also a suitable context in which to consider an unpublished definition of a general integral recently introduced by N. Bourbaki.[2] As we shall see below, the general integral of Bourbaki differs in few of its essential properties from the one treated in these notes. By comparing the two integrals we obtain solutions to a number of problems raised by Theorem 18 of our second note and mentioned in the corresponding footnote.

On a fixed abstract set $X$, let $\mathfrak{E}'$ and $\mathfrak{E}''$ be families of elementary functions; and let $E'$ and $E''$ be elementary integrations defined over the respective families. It is easy to verify that the contractions of $E'$ and $E''$ to the family $\mathfrak{E} = \mathfrak{E}' \cap \mathfrak{E}''$ share with them the postulated properties set out in I (1), I (2). Consequently the study of the relations between $E'$ and $E''$ breaks naturally into two parts—the first consisting in the study of two elementary integrals defined over the same family of elementary functions (viz., the contractions of $E'$ and $E''$ to $\mathfrak{E}$); the second in the study of two elementary integrals, one of which is an extension or contraction of the other (viz., $E'$ or $E''$ and its contraction to $\mathfrak{E}$). We shall discuss these situations separately below.

As linear functionals the elementary integrals over a fixed family $\mathfrak{E}$ of elementary functions constitute an ordered additive semigroup admitting multiplication by positive real scalars: the definitions of $\alpha E$, $E' + E''$, and the ordering relation $\leqq$ are given by requiring that $(\alpha E)(f) = \alpha(E(f))$, $(E' + E'')(f) = E'(f) + E''(f)$, $E'(f) \leqq E''(f)$ for all $f$ in $\mathfrak{E}$; and the various needed properties, including those demanded in I (1) and I (2), can be verified without difficulty. We note in particular that if $E = E' + E''$ then $E' \leqq E$ and $E'' \leqq E$ and that if $E' \leqq E$ then there is a unique $E''$, given by $E''(f) = E(f) - E'(f)$, such that $E = E' + E''$. In order to examine the relations between given elementary integrals $E'$ and $E''$, we shall first consider how each is related to the sum $E = E' + E''$ and then combine such information as we obtain in this way to yield results on the main problem. The various quantities and other mathematical objects associated with the elementary integrals $E$, $E'$ and $E''$ will be distinguished by the use of the corresponding primes. Most of our results are of such a nature that we shall confine attention to the case where $1 \in \mathfrak{M}$, that is, where the constant function everywhere equal to 1 is measurable in the theory associated with the elementary integral $E$. For practical purposes, indeed, there would be no objection to assuming that the property I (3), which implies $1 \in \mathfrak{M}$, holds for $\mathfrak{E}$. All our deeper results depend upon the additional assumption that

(1) *$\mathfrak{L}$ contains a sequence $\{f_n\}$ such that $\{x;\ \sup_n |f_n(x)| = 0\}$ is a null set.*[3]

An example given by Saks[4] shows that this assumption, or a similar one, is indispensable. Now we may note without formal proof that

(2) *the relation $E' \leqq E$ on $\mathfrak{E}$ implies $N' \leqq N$, $\mathfrak{F}' \supset \mathfrak{F}$, $\mathfrak{L}' \supset \mathfrak{L}$, and $\mathfrak{M}' \supset \mathfrak{M}$; it also implies that $L' \leqq L$ in $\mathfrak{L}$.*

These preliminaries settled, we proceed to discuss an important portion of the Radon-Nikodym theorem,[4] modifying for our purposes a method of proof originally introduced by J. v. Neumann:[5]

(3) (Radon-Nikodym) *if $E' \leqq E$, there exists an essentially unique function $\phi'$ in $\mathfrak{M}$, $0 \leqq \phi' \leqq 1$, such that $f \in \mathfrak{L}'$ if and only if $\phi' f \in \mathfrak{L}$, while $L'(f) = L(\phi' f)$; in particular, $E'(f) = L(\phi' f)$ when $f \in \mathfrak{E}$.*

For $h$ and $k$ in $\mathfrak{L}_2$ the product $hk$ is in $\mathfrak{L}$. Hence $L'(hk)$ is a bilinear functional defined for all $h$ and $k$ in $\mathfrak{L}_2$. Since $|L'(hk)| \leqq L'(|hk|) \leqq L(|hk|) \leqq L(h^2)^{1/2}L(k^2)^{1/2}$, a standard theorem[6] about generalized Hilbert spaces asserts the existence of a continuous linear transformation $A$ of $\mathfrak{L}_2$ into itself such that $L'(hk) = L((Ah)k)$. If $g$ is any bounded member of $\mathfrak{M}$, then $gh$ and $gk$ are in $\mathfrak{L}_2$. The equations $L((A(gh))k) = L'((gh)k) = L'(h(gk)) = L((Ah)gk) = L((g(Ah))k)$ imply that $A(gh) = gAh$ almost everywhere. In order to exploit the latter relation, we use (1) to construct a bounded function $g_0$ in $\mathfrak{L}_2$ such that $\{x;\ g_0(x) = 0\}$ is a null set. Supposing (as we obviously may) that the sequence $\{f_n\}$ of (1) contains no null function, we put $f_0 = \sum_{n=1}^{\infty} 2^{-n}|f_n|/L(|f_n|) \in \mathfrak{L}$ in harmony with I (12) and observe that $\{x;\ f_0(x) = 0\}$ is a null set. We can then put $g_0 = (\min\ (1, f_0))^{1/2}$. For any $h$ in $\mathfrak{L}_2$ the function $h_n = \text{mid}\ (h, n, -n)$ is bounded and converges in $\mathfrak{L}_2$ to $h$ as $n \to \infty$. Thus $g_0(Ah_n) = A(g_0 h_n) = A(h_n g_0) = h_n(Ag_0) = (Ag_0)h_n$ and passage to the limit in the extreme terms yields $g_0(Ah) = (Ag_0)h$ almost everywhere. A measurable function $\phi'$ is defined by putting $\phi' = Ag_0/g_0$ almost everywhere. Evidently $Ah = \phi' h$. Hence if $f \in \mathfrak{L}$ and $h = |f|^{1/2} \in \mathfrak{L}_2$, $k = |f|^{1/2}\ \text{sgn}\, f \in \mathfrak{L}_2$, we have $L'(f) = L'(hk) = L(\phi' hk) = L(\phi' f)$. The inequalities $0 \leqq L'(f) \leqq L(f)$ hold whenever $f \geqq 0$ and imply with the help of (1) that $0 \leqq \phi' \leqq 1$ almost everywhere. Similarly, if $\phi_1'$ and $\phi_2'$ are two measurable functions such that $L(\phi_1'f) = L'(f) = L(\phi_2'f)$ for all $f$ in $\mathfrak{L}'$, we have $L((\phi_1' - \phi_2')f) = 0$ for all such $f$ and hence $\phi_1' = \phi_2'$ almost everywhere. *Thus $\phi'$ is uniquely determined (in the obvious sense).* In proving the remainder of the theorem, we make use of the definition of $\mathfrak{L}'$. A function $f$ is in $\mathfrak{L}'$ if and only if to an arbitrary $\epsilon > 0$ there correspond functions $g$ and $f_n$ in $\mathfrak{E}$ such that $|f - g| \leqq \sum_{n=1}^{\infty} |f_n|$, $\sum_{n=1}^{\infty} E'(|f_n|) \leqq \epsilon$. Clearly we have $|L'(f) - L'(g)| \leqq \epsilon$ under these circumstances. Since $|\phi'f - \phi'g| \leqq \sum_{n=1}^{\infty} \phi'|f_n|$, $\phi'g \epsilon \mathfrak{L}$ we see that $N(\phi'f - \phi'g) \leqq \sum_{n=1}^{\infty} N(\phi'|f_n|) = \sum_{n=1}^{\infty} L(\phi'|f_n|) = \sum_{n=1}^{\infty} L'(|f_n|) = \sum_{n=1}^{\infty} E'(|f_n|) \leqq \epsilon$ and hence that $\phi'f \epsilon \mathfrak{L}$, $|L(\phi'f) - L(\phi'g)| \leqq \epsilon$. The equation $L(\phi'g) = L'(g)$ yields $|L'(f) - L(\phi'f)| \leqq 2\epsilon$. In considering the converse, it is convenient to observe that the relation $N' \leqq N$ requires every null set in the $E$-theory to be a null set in the $E'$-theory also. Thus when $\phi'f \epsilon \mathfrak{L}$ we see that $f$ is finite almost everywhere in the sense of either theory. Since we can examine separately the two functions $|f|$ and $|f| - f$, where $\phi'|f| = |\phi'f| \epsilon \mathfrak{L}$, $\phi'(|f| - f) = \phi'|f| - \phi'f \epsilon \mathfrak{L}$ and $f = |f| - (|f| - f)$, there is no loss of generality in supposing that $f \geqq 0$. Then on putting $g_n = f/(n\phi' + 1)$, $h_n = n\phi'f/(n\phi' + 1)$ we see that the sequences $\{g_n\}$, $\{h_n\}$ are respectively monotonically decreasing and increasing almost everywhere to the corresponding limit functions $g$ and $h$. Since $g_n + h_n = f$ we have $g + h = f$, almost everywhere; and we verify similarly that $\phi'g = 0$, $\phi'h = \phi'f$ almost everywhere. The relation $h_n \epsilon \mathfrak{L}$ implies that $h_n \epsilon \mathfrak{L}'$ and $L'(h_n) = L(\phi'h_n) \leqq L(\phi'f)$. Hence I (12) shows that $h \epsilon \mathfrak{L}'$. Turning now to a consideration of $g$, we shall show that $g \epsilon \mathfrak{L}'$ and $L'(g) = 0$. For this purpose, we construct a function $f_0 \geqq 0$ in $\mathfrak{L}$ such that $\{x; f_0(x) = 0\}$ is a null set, in the manner indicated above. Since $f_0 \epsilon \mathfrak{L}'$ and $\phi'f_0 \epsilon \mathfrak{L}$ we can express $f_0$ also in the form $g_0 + h_0$ just described. We note that $g_0 = f_0 - h_0 \epsilon \mathfrak{L}'$, $N'(g_0) = L'(g_0) = L(\phi'g_0) = 0$. Clearly we have $g_0(x) > 0$ whenever $\phi'(x) = 0$, except possibly on a null set; and hence we also have $g_0(x) > 0$ whenever $g(x) > 0$, with a like qualification. The relation $g \leqq g_0 + g_0 + g_0 + \ldots$ therefore holds almost everywhere and implies that $N'(g) \leqq N'(g_0) + N'(g_0) + N'(g_0) + \ldots = 0$. Thus $g$ is a null function in the $E'$-theory and as such has the properties desired. Since $g$ and $h$ are both in $\mathfrak{L}'$ we have $f = g + h \epsilon \mathfrak{L}'$. This completes the proof of (3). It is desirable to mention without proof that (3) has the following easy converse:

(4) *if $E$ is given and $\phi$ is any function in $\mathfrak{M}$ such that $0 \leqq \phi \leqq 1$, then the operation $E'$ defined over $\mathfrak{E}$ by the equation $E'(f) = L(\phi f)$ is an elementary integration such that $E' \leqq E$; and the function $\phi'$ obtained by virtue of (3) differs from $\phi$ at most on a null set.*

Clearly (3) and (4) have the following interpretation: the correspondence $E' \rightarrow \phi'$ is an isomorphism (both algebraic and ordinal) between the totality of $E' \leqq E$ and the totality of $\phi' \in \mathfrak{M}$, $0 \leqq \phi' \leqq 1$, identified modulo null functions. In particular, the familiar lattice properties of this totality of functions are reflected in the system of all $E' \leqq E$. It is then easy to see that the system of all elementary integrations is a lattice with respect to the ordering relation $\leqq$; and that in it every monotonely decreasing sequence has a greatest lower bound.[7] The isomorphism leads directly to the complete Radon-Nikodym theorem and to the Lebesgue decomposition theorem, as we shall now show. It is convenient to introduce the following definition of a new order relation $\mapsto$ between elementary integrations: $E' \mapsto E''$ if and only if every null set in the $E''$-theory is a null set in the $E'$-theory (or, equivalently, if and only if every null function in the $E''$-theory is a null function in the $E'$-theory). Evidently $E' \leqq E''$ implies $E' \mapsto E''$; and when $E = E' + E''$ the condition on the associated functions $\phi'$ and $\phi''$ equivalent to the condition $E' \mapsto E''$ is that $\phi''(x) = 0$ imply $\phi'(x) = 0$ for almost all $x$. We can now state the Lebesgue decomposition theorem as follows:[8]

(5) (Lebesgue) *if $E'$ and $E''$ are given elementary integrations over $\mathfrak{E}$, then $E'$ has a unique decomposition of the form $E' = E'_1 + E'_2$ where $E'_1 \mapsto E''$ and $(E'_2 \cap E'')(|f|) = 0$ for all $f$ in $\mathfrak{E}$.*

The proof is given by putting $E = E' + E''$ and considering the associated functions $\phi'$ and $\phi''$. In order to obtain the desired decomposition we evidently have to determine $\phi'_1$ and $\phi'_2$ so that $\phi' = \phi'_1 + \phi'_2$, $\phi''(x) = 0$ implies $\phi'_1(x) = 0$ for almost all $x$, and $\min(\phi'_2, \phi'') = 0$ almost everywhere. Then $\phi'_1$ and $\phi'_2$ are essentially unique and can be expressed in terms of $\phi'$ and $\phi''$ by the formulae $\phi'_1 = \lim_{n\to\infty} n\phi''\phi'/(n\phi'' + 1)$, $\phi'_2 = \lim_{n\to\infty} \phi'/(n\phi'' + 1)$ similar to ones used in the discussion of (3). The complete Radon-Nikodym theorem reads[9]

(6) (Radon-Nikodym) *if $E'$ and $E''$ are given elementary integrations over $\mathfrak{E}$ such that $E' \mapsto E''$, then there exists an essentially unique non-negative function $\chi$ in $\mathfrak{M}''$ such that $f \in \mathfrak{L}'$ if and only if $\chi f \in \mathfrak{L}''$, while $L'(f) = L''(\chi f)$; in particular $E'(f) = L''(\chi f)$ when $f \in \mathfrak{E}$. Conversely, if $E''$ is given and if $\chi$ is any non-negative function in $\mathfrak{M}''$ such that $f \in \mathfrak{E}$ implies $\chi f \in \mathfrak{L}''$, then the operation $E'$ defined by the equation $E'(f) = L''(\chi f)$ is an elementary integration such that $E' \mapsto E''$.*

The proof is again based on the isomorphism established in (3) and (4). For given $E'$ and $E''$, we put $E = E' + E''$ and determine the associated functions $\phi'$ and $\phi''$ in $\mathfrak{M} \subset \mathfrak{M}' \cap \mathfrak{M}''$ in accordance with (3). Since $\phi' + \phi'' = 1$ and $\phi''(x) = 0$ implies $\phi'(x) = 0$ for almost all $x$, we see that $\phi''$ vanishes at most on a null set. Consequently the function $\chi = \lim_{n\to\infty} n\phi'/$

$(n\phi'' + 1)$ is equal almost everywhere to $\phi'/\phi''$; and the equations $\phi' = \chi/(\chi + 1)$, $\phi'' = 1/(\chi + 1)$ hold almost everywhere. It is clear that $\chi \in \mathfrak{M} \subset \mathfrak{M}'$. From (3) we now conclude that $f \in \mathfrak{L}'$ if and only if $\phi''(\chi f) = \phi' f \in \mathfrak{L}$ and hence if and only if $\chi f \in \mathfrak{L}''$. When these conditions are verified we have $L'(f) = L(\phi' f) = L(\phi''(\chi f)) = L''(\chi f)$. On the other hand, when $E''$ and $\chi$ are given it is easy to verify that $E'$ is an elementary integration. We therefore put $E = E' + E''$ and determine the associated functions $\phi'$ and $\phi''$ in $\mathfrak{M}$. If $f \in \mathfrak{E}$ we have $L(\phi' f) = E'(f) = L''(\chi f) = L(\phi''\chi f)$. Thus $\chi$ is finite for almost all $x$ and the product $\phi''\chi$ is defined and non-negative almost everywhere. We observe now that the equation $L(\phi' f) = L(\phi''\chi f)$ can be established by continuity for all $f$ in $\mathfrak{L}$. For such an $f$, let $\{f_n\}$ be a sequence of functions in $\mathfrak{E}$ which converges in $\mathfrak{L}$ to $f$. We may suppose in addition that $\lim\limits_{n\to\infty} f_n$ exists almost everywhere and differs from $f$ only on a null set. Then $\lim\limits_{n\to\infty} \phi''\chi f_n = \phi''\chi f$ almost everywhere. At the same time $L(|\phi''\chi f_m - \phi''\chi f_n|) = L(\phi''\chi|f_m - f_n|) = L(\phi'|f_m - f_n|) \leqq L(|f_m - f_n|)$ so that $\{\phi''\chi f_n\}$ converges in $\mathfrak{L}$—necessarily to $\phi''\chi f$. We obviously have $L(\phi''\chi f) = \lim\limits_{n\to\infty} L(\phi''\chi f_n) = \lim\limits_{n\to\infty} L(\phi' f_n) = L(\phi' f)$. By (1) we can construct the function $f_0$ used in the proof of (3). Since we have $\phi''\chi f_0 \in \mathfrak{L}$ it follows that $\phi''\chi = \lim\limits_{n\to\infty} n\phi''\chi f_0/(nf_0 + 1) \in \mathfrak{M}$. With further use of (1) we can infer from the fact that $L(\phi' - \phi''\chi)f) = 0$ for all $f$ in $\mathfrak{L}$, $\phi' - \phi''\chi \in \mathfrak{M}$, that $\phi' - \phi''\chi = 0$ almost everywhere. Thus $\phi''(x) = 0$ implies $\phi'(x) = 0$ for almost all $x$, and we must have $E' \rightarrowtail E''$. The formulae connecting $\phi'$, $\phi''$, $\chi$ are evidently the same as those noted above, and $\chi \in \mathfrak{M}$. The relations $E'' \rightarrowtail E$, $E \rightarrowtail E''$ are both valid. We conclude the work of this section by an observation concerning the equivalence relation $\sim$ defined by putting $E' \sim E''$ if and only if both $E' \rightarrowtail E''$ and $E'' \rightarrowtail E'$. We have:[10]

(7) *with respect to the order relation* $\rightarrowtail$ *and the equivalence relation* $\sim$ *the totality of elementary integrations defined on the family* $\mathfrak{E}$ *of elementary functions is a* $\sigma$*-additive generalized Boolean algebra.*

If we consider the $E'$ such that $E' \leqq E$ and interpret the relations $\sim$ and $\rightarrowtail$ in the isomorphic system of functions $\phi'$, we see at once that each equivalence class contains exactly one characteristic function (modulo null functions) and that for characteristic functions the relation corresponding to $\rightarrowtail$ is the natural ordering for real functions. The system of elementary integrations $E'$ such that $E' \leqq E$ is therefore a $\sigma$-additive Boolean algebra, as stated. The remainder of the theorem then follows, provided that the $\sigma$-additivity be understood to mean that any *bounded* sequence has a Boolean sum.

Turning now to the second part of our problem—the study of the extensions of a given $E$—we see at once that

(8) *if* $E'$ *is an extension of* $E$ *from* $\mathfrak{E}$ *to* $\mathfrak{E}'$, *then* $N' \leqq N$, $\mathfrak{F}' \supset \mathfrak{F}$, $\mathfrak{L}' \supset \mathfrak{L}$, *and* $L'$ *is an extension of* $L$ *from* $\mathfrak{L}$ *to* $\mathfrak{L}'$; *if in particular* $\mathfrak{E} \subset \mathfrak{E}' \subset \mathfrak{L}$ *then* $N' = N$, $\mathfrak{F}' = \mathfrak{F}$, $\mathfrak{L}' = \mathfrak{L}$ *and* $L' = L$.

The first part is evident, the second almost so. However we may remark that the second part results most easily from considering $L$ as an extension $E''$ of $E$ and $E'$ over $\mathfrak{E}'' = \mathfrak{L}$, since II (14) then immediately gives $N'' = N$ and hence $N \leqq N'' \leqq N' \leqq N$, $N' = N$. It appears that very little indeed can be said about the general problem of extensions, largely because the condition of I (2) makes the construction of extensions a matter of considerable delicacy. When this condition is dropped it is possible to obtain certain results which are not altogether without interest. We may illustrate this by the following theorem:[11]

(9) *the elementary integration* $E$ *over* $\mathfrak{E}$ *has an extension* $J$ *over* $\mathfrak{F}$ *which is a positive linear functional with the properties* $0 \leqq J(|f|) \leqq N(f)$, $J(f) = L(f)$ *when* $f \in \mathfrak{L}$.

The proof is an immediate consequence of the Hahn-Banach theorem: we put $p(f) = N(f^+)$ where $f^+ = \frac{1}{2}(|f| - f)$, observing that $0 \leqq p(f)$, $p(f + g) \leqq p(f) + p(g)$, $p(\alpha f) = \alpha p(f)$ when $\alpha \geqq 0$, and $L(f) \leqq L(f^+) = p(f)$ when $f \in \mathfrak{L}$; and we conclude that $L$ has a linear extension $J$ over $\mathfrak{F}$ such that $J(f) \leqq p(f)$ for all $f$ in $\mathfrak{F}$, a condition which implies for any $f \geqq 0$ that $J(-f) \leqq p(-f) = N((-f)^+) = N(0) = 0$, $J(f) \geqq 0$. The relations $0 \leqq J(|f|) \leqq p(|f|) = N(|f|^+) = N(|f|) = N(f)$ then hold also.

Having done what little we can with the extension problem, we shall now turn to a discussion of the general integral due to Bourbaki. This integral is defined in a manner quite similar to that followed in these notes; but somewhat more stringent conditions are imposed on the elementary functions and integrals, while more general processes are employed for the construction of the general integral. As we shall see, it follows that whenever the Bourbaki theory is applicable our theory is also and yields a general integration which has the general integration of Bourbaki as an extension, of remarkably simple type: this extension can, in fact, be obtained essentially by augmenting the class of null functions. It is thus apparent that the discussion of the Bourbaki integral is appropriately included in the present note. The definition of the Bourbaki integral rests upon the use of function-systems filtering with respect to the natural ordering for functions: a class of functions is said to be a filtering (or directed) system if whenever it contains $f$ and $g$ it also contains a function $h$ such that $h \geqq \max(f, g)$. Here we shall always denote such a system by the letter $\mathfrak{K}$. Let there be given a system $\mathfrak{E}$ of elementary functions and an elementary integration $E$ which satisfy the conditions of I (1) and of the following strengthened version of I (2), namely:

($\bar{2}$) *if* $\mathfrak{K}$ *is a filtering system in* $\mathfrak{E}^+$, *the system of non-negative functions in* $\mathfrak{E}$, *and if* $f$ *is any function in* $\mathfrak{E}$ *such that* $|f(x)| \leqq \sup_{k \in \mathfrak{K}} k(x)$, *then* $E(|f|) \leqq \sup_{k \in \mathfrak{K}} E(k)$.

A quantity analogous to $N(f)$ can then be introduced through the definition

$$(\bar{4})\quad \bar{N}(f) = \inf_{\mathfrak{K} \subset \mathfrak{E}^+} \{\lambda;\ \lambda = \sup_{k \in \mathfrak{K}} E(k),\ |f(x)| \leqq \sup_{k \in \mathfrak{K}} k(x)\}$$

If $|f| \leqq \sum_{n=1}^{\infty} |f_n|, f_n \in \mathfrak{E}$, we can take $\mathfrak{K}$ as the totality of partial sums of this infinite series; and we immediately infer that $\bar{N} \leqq N$. Moreover properties I (5)–I (9) are readily verified for $\bar{N}$. Hence whatever parts of the theory presented in these notes have been built directly on these properties have valid analogs in the theory of Bourbaki. In particular, we see that $\bar{\mathfrak{F}} = \{f;\ \bar{N}(f) < +\infty\}$ is a Banach space, and we can define $\bar{\mathfrak{L}}$ as the closure of $\mathfrak{E}$ in $\bar{\mathfrak{F}}$, taking $\bar{L}(f) = \bar{N}(f^+) - \bar{N}(f^-) = \bar{F}(f)$ for $f \in \bar{\mathfrak{L}}$. A detailed check shows that the only results in our theory which depend upon a direct use of the definition of $N$ are I (5)–I (9), II (10), II (14), III (1) and one observation in the proof of (3) above. Thus it remains for us to prove analogs of the last-mentioned results for the Bourbaki theory. We must observe further that in connection with the Fubini theorem in III and also at certain points in the present note it was necessary to check the validity of I (2) for particular families of functions. At the same points the corresponding check of ($\bar{2}$) clearly has to be made in the Bourbaki theory. This can be accomplished in every instance without difficulty. Our program for treating the Bourbaki integral has therefore been reduced to two steps, the first being to prove a result of general interest from which II (14), II (10) and the point referred to in the proof of (3) above, all easily follow, namely:

(10) $f \in \bar{\mathfrak{F}}$ *if and only if* $f = g + h$ *where* $g \in \mathfrak{F}$ *and* $h$ *is a null function in the sense that* $\bar{N}(h) = 0$, *it being possible when* $f \geqq 0$ *to choose* $g \geqq 0$ *and* $h \geqq 0$ *so that* $\bar{N}(f) = \bar{N}(g) = N(g)$; *likewise,* $f \in \bar{\mathfrak{L}}$ *if and only if* $f = g + h$ *where* $g \in \mathfrak{L}$ *and* $\bar{N}(h) = 0$—*and, when these conditions hold,* $\bar{L}(f) = \bar{L}(g) = L(g)$.

We first observe that in consequence of the relation $N \leqq \bar{N}$ we have $\bar{\mathfrak{F}} \supset \mathfrak{F}$, $\bar{\mathfrak{L}} \supset \mathfrak{L}$, and $\bar{L}(f) = L(f)$ for all $f \in \mathfrak{L}$. Thus $\bar{L}$ is an extension of $L$. The sufficiency of the conditions stated in the theorem is evident. In discussing their necessity, we may treat the case where $f \geqq 0$ since the general case can be reduced to it by consideration of the functions $f^+$ and $f^-$. Assuming $f \geqq 0$

and $\bar{N}(f) < +\infty$, we choose a filtering system $\mathfrak{K}_p \subset \mathfrak{E}^+$ so that $0 \leqq f(x) \leqq \sup_{k \in \mathfrak{K}_p} k(x)$, $\bar{N}(f) \leqq \lambda_p = \sup_{k \in \mathfrak{K}_p} E(k) \leqq \bar{N}(f) + 1/p$. Evidently this system contains an increasing sequence $\{k_{np}\}$ such that $\lambda_p = \lim_{n \to \infty} E(k_{np})$. In accordance with I (12) this sequence has a limit $k_p$ in $\mathfrak{L}$. Putting $g_p = \min(f, k_p) \geqq 0$, we see that $g_p \in \mathfrak{F}$. We also put $h_p = \max(f, k_p) - k_p \geqq 0$ and show that $\bar{N}(h_p) = 0$. The equation $f = g_p + h_p$ then gives the desired result. In order to estimate $\bar{N}(h_p)$ we put $h_{np} = \max(f, k_{np}) - k_{np}$, obtaining a sequence which decreases with $1/n$ to the limit $h_p$. Hence $\bar{N}(h_p) \leqq \bar{N}(h_{np})$. The functions $\max(k, k_{np}) - k_{np} \geqq 0$, $k \in \mathfrak{K}_p$, constitute a filtering system in $\mathfrak{E}^+$ and for each $x$ in $X$ have a supremum not less than $h_{np}(x) \geqq 0$. Consequently $\bar{N}(h_{np}) \leqq \sup_{k \in \mathfrak{K}_p} (E(\max(k, k_{np}) - E(k_{np})) = \lambda_p - E(k_{np})$. Passing to the limit we have $\bar{N}(h_{np}) \to 0$ and hence $\bar{N}(h_p) = 0$. It is clear that $\bar{N}(g_p) \leqq \bar{N}(f) \leqq \bar{N}(g_p) + \bar{N}(h_p)$ and hence that $\bar{N}(g_p) = \bar{N}(f)$. However all we know concerning $N(g_p)$ is that $\bar{N}(f) = \bar{N}(g_p) \leqq N(g_p) \leqq N(k_p) = \lambda_p \leqq \bar{N}(f) + 1/p$. To complete the discussion we therefore put $g = \inf g_p \geqq 0$, noting that $f \geqq g$, $g \in \mathfrak{F}$, $f = g$ almost everywhere in the Bourbaki theory. Writing $h = f - g \geqq 0$, we have $\bar{N}(h) = 0$. Now $\bar{N}(f) \leqq \bar{N}(g) + \bar{N}(h) = \bar{N}(g) \leqq N(g) \leqq N(g_p) \leqq \bar{N}(f) + 1/p$ for all $p$, and $\bar{N}(f) = \bar{N}(g) = N(g)$, as we wished to prove. Assuming now that $f$ is an arbitrary function in $\mathfrak{L}$ we know that there is a sequence $\{f_n\}$ in $\mathfrak{E}$ such that $\bar{N}(f - f_n) = \bar{L}(f - f_n) \to 0$ while $\bar{N}(f_n - f_{n+1}) = \bar{L}(|f_n - f_{n+1}|) = E(|f_n - f_{n+1}|) \leqq 2^{-n}$. It follows that $\{f_n\}$ also converges in $\mathfrak{L}$ to a function $g$. Clearly we must have $f = g$ almost everywhere in the Bourbaki theory, since both functions are equal almost everywhere to $\limsup_{n \to \infty} f_n$. Thus we have $f = g + h$ where $\bar{N}(h) = 0$. It is easy to see that $\bar{L}(f) = \bar{L}(g) + \bar{L}(h) = \bar{L}(g) = L(g)$ here. This completes the proof. The second step of our program is to give a direct demonstration of III (1) for the Bourbaki theory. We have

(11) *if the elementary integrations* $E_x$, $E_y$ *and* $E_z$ *satisfy the relation* $E_z \subset E_x * E_y$, *then the corresponding operations* $\bar{N}_x$, $\bar{N}_y$, $\bar{N}_z$ *are such that* $\bar{N}_z f(z) \geqq \bar{N}_x \bar{N}_y f(x, y)$ *for every* $f$ *in* $\mathfrak{G}(z)$.

The proof is simple. Since there is nothing to prove unless $\bar{N}_z f(z) < +\infty$, we assume the latter relation. We can then choose for any given $\epsilon > 0$ a filtering system $\mathfrak{K} \subset \mathfrak{E}^+(z)$ such that $|f(z)| \leqq \sup_{k \in \mathfrak{K}} k(z)$, $\bar{N}_z f(z) \leqq \sup_{k \in \mathfrak{K}} E(k) \leqq \bar{N}_z f(z) + \epsilon$. The functions $k(x, y)$ for fixed $x$ constitute a filtering system in $\mathfrak{E}^+(Y)$ and their supremum is not less than $|f(x, y)|$. Hence we must have $\bar{N}_y f(x, y) \leqq \sup_{k \in \mathfrak{K}} E_y k(x, y)$. Since the functions $E_y k(x, y)$, $k \in \mathfrak{K}$, constitute a filtering system in $\mathfrak{E}^+(X)$, we must have $\bar{N}_x \bar{N}_y f(x, y) \leqq \sup_{k \in \mathfrak{K}} E_x E_y k(x, y) = \sup_{k \in \mathfrak{K}} E_z k(z) \leqq \bar{N}_z f(z) + \epsilon$. The theorem then follows at once. We now see that all the results established in our theory hold, mutatis mutandis, for the theory of Bourbaki. The latter theory is calculated to give the best possible handling of the case where $\mathfrak{C}$ is the family of all continuous functions with compact nuclei on a locally compact space $X$. We shall therefore conclude our remarks with some comments on this case. In the Bourbaki theory it is not difficult to verify that the Riesz-Markoff-Kakutani theorem, II (18), assumes precisely the form announced by Markoff: the measure is a regular measure such that every compact subset of $X$ is measurable. Thus where our theory produces a Baire measure, the Bourbaki theory produces a Borel measure. When $X$ is the homogeneous space of the left cosets of a closed subgroup $Y$ of a locally compact topological group $Z$, the discussion given in III can be repeated verbatim in the Bourbaki theory. Now, however, the final result is a generalization of Ambrose's theorem cited there. We thus obtain a simplified proof in which we do not need to assume, as Ambrose does, that $Y$ is a normal subgroup.

---

[1] Stone, M. H., "Notes on Integration, I, II, III," these PROCEEDINGS, **34**, 336–342, 447–455, 483–490 (1948); cited here as I, II, III, respectively.

[2] After the preparation of our third note, we learned that Bourbaki's projected treatment of the theory of integration had assumed a form remarkably like that described in these notes. The resemblance extends even to some of the finer details. The comparative study outlined in the present note should suffice to clarify the technical relations between the two theories.

[3] If such a sequence exists it can be replaced by a similar sequence in $\mathfrak{C}$, by a simple application of the definition of $\mathfrak{L}$.

[4] Saks, S., *Theory of the Integral*, 2nd revised ed., Warszawa-Lwow, 1937, pp. 32–36.

[5] v. Neumann, J., "On Rings of Operators, III," *Ann. Math.*, **41**, 94–161 (1940), especially pp. 127–129.

[6] Stone, M. H., *Linear Transformations in Hilbert Space and Their Applications to Analysis*, New York, 1932, Theorem 2.28. The proof given there for the complex separable case applies also to the real unrestricted case, which we need here.

[7] Riesz, F., "Sur quelques notions fondamentales dans la théorie générale des opérations linéaires," *Ann. Math.*, **41**, 174–206 (1940), has shown how the lattice properties can be obtained directly. Stronger properties than those stated here are actually valid and can be established in general by Riesz's methods or in this instance by direct examination of the function-lattice.

[8] Saks, S., *loc. cit.*[4]

[9] Saks, S., *loc. cit.*[4]

[10] An important application of (7) has been made in Plesner, A. I., and Rohlin, V. A., "Spectral Theory of Linear Operators," *Uspehi Matem. Nauk*, N.S. 1 (11), no. 1, 71–191 (1946) [Russian], §§ 22–27. See the writer's review in *Math. Rev.*, **9**, 43 (1948).

[11] This result is to be compared with one given by Banach, S., *Théorie des Opérations Linéaires*, Warszawa, 1932, pp. 29–32.

# ON THE DEGREE OF CONVERGENCE OF HARMONIC POLYNOMIALS TO HARMONIC FUNCTIONS

BY J. L. WALSH, W. E. SEWELL AND H. M. ELLIOTT*

HARVARD UNIVERSITY, U. S. ARMY, AND HARVARD UNIVERSITY

Communicated December 3, 1948

The purpose of this note is to state without proof certain results on polynomial approximation to harmonic functions whose proofs will appear elsewhere later.

In the plane of the complex variable $z = x + iy$, a (real) function $u(z)$ harmonic in the interior of the Jordan curve $C$ and continuous in the corresponding closed region $\bar{C}$ can be represented in $\bar{C}$ by a uniformly convergent series of harmonic polynomials.[1] If $C_R$ denotes the image of the circle $|w| = R > 1$ under the conformal map of the exterior of $C$ onto the region $|w| > 1$ of the $w$-plane so that the points at infinity in the two planes correspond to each other, and if $u(z)$ is harmonic throughout the interior of $C_R$, then there exist harmonic polynomials $p_n(z)$ of respective degrees $n = 1, 2, \ldots,$ such that

$$\limsup_{n \to \infty} [\max. |u(z) - p_n(z)| \text{ on } C]^{1/n} \leqq 1/R; \qquad (1)$$

conversely if harmonic polynomials $p_n(z)$ of respective degrees $n$ exist so that (1) is satisfied, the sequence $p_n(z)$ converges throughout the interior of $C_R$, and the function $u(z)$ if suitably defined exterior to $C$ is consequently harmonic throughout the interior of $C_R$. We sharpen these prior results[2] by studying now the more delicate problem of relating continuity properties of $u(z)$ (in the form of Lipschitz[3] conditions of arbitrary order on the function or its derivatives) on $C$ or $C_R$ to degree of convergence of the sequence $p_n(z)$ in $\bar{C}$.

This theory is closely related to the analogous theory for approximation of analytic functions by polynomials in the complex variable.[4] But the present results are by no means merely an obvious application of the earlier ones.

We define four classes of functions as follows. Let $C$ be a rectifiable Jordan curve. We say that the real function $u(z)$ belongs to the class $L(k, \alpha)$, symbolically $u(z) \in L(k, \alpha)$, $k \geqq 0$, on $C$ if $u(z)$ is harmonic interior to $C$ and continuous on $\bar{C}$; if the derivative $\partial^k u(z)/\partial s^k$ exists on $C$, where $s$ is arc length measured along $C$; and if $\partial^k u(z)/\partial s^k$ satisfies on $C$ a Lipschitz condition of order $\alpha$, $0 < \alpha \leqq 1$. We say $u(z) \in \text{Log}\ (k, 1)$ on $C$ if $u(z)$ is harmonic interior to $C$, continuous on $\bar{C}$, and $\partial^k u/\partial s^k$ exists on $C$ and satisfies the condition

$$\left| \frac{\partial^k u(z_1)}{\partial s^k} - \frac{\partial^k u(z_2)}{\partial s^k} \right| \leqq L|z_1 - z_2|\ \left|\log |z_1 - z_2|\right|, \ z_1, z_2 \text{ on } C, \qquad (2)$$

where $|z_1 - z_2|$ is sufficiently small.

The complex-valued function $f(z)$ is said to belong to the class $L_A(k, \alpha)$ on the closed region $\bar{C}$ if $f(z)$ is analytic interior to $C$, continuous on $\bar{C}$, and if $f^{(k)}(z)$ exists and satisfies a Lipschitz condition of order $\alpha$ on $\bar{C}$. Finally, $f(z) \in \text{Log}_A\ (k, 1)$ on $\bar{C}$ if $f(z)$ is analytic interior to $C$, continuous on $\bar{C}$, and if $f^{(k)}(z)$ exists and satisfies a condition of the form (2) on $\bar{C}$.

If $C$ consists of a finite number of mutually exterior rectifiable Jordan curves, we say $u(z) \in L(k, \alpha)$ on $C$ if $u(z) \in L(k, \alpha)$ on each component of $C$; we make analogous definitions for the other three classes.

THEOREM 1. *Let $C$ consist of a finite number of mutually exterior analytic Jordan curves. Let $u(z) \in L(k, \alpha)$ on $C$, $k \geqq 0$, $0 < \alpha \leqq 1$. Then for each $n$, $n = 1, 2, \ldots$, there exists a harmonic polynomial $p_n(z)$ such that*

$$|u(z) - p_n(z)| \leqq M/n^{k+\alpha}, \qquad z \text{ on } \bar{C}.$$

We say that a curve $C$ is of Type D if (i) $C$ is a rectifiable Jordan curve whose arc and chord are uniformly infinitesimals of the same order, and (ii) there exists a number $\delta_0 > 0$ such that if $P$ is any point of $C$, there is a circle of radius $\delta_0$ through $P$ whose closed interior lies in $\bar{C}$.

As a converse to Theorem 1 we have

THEOREM 2. *Let $C$ be a curve of Type D, and let $u(z)$ be defined in $\bar{C}$. For each $n$, $n = 1, 2, \ldots$, let a harmonic polynomial $p_n(z)$ exist such that*

$$|u(z) - p_n(z)| \leqq M/n^{k+\alpha}, \qquad z \text{ in } \bar{C},\ k \geqq 0.$$

*Then $u(z) \in L(0, \alpha)$ on $C$ for $0 < \alpha < 1$ and $u(z) \in \text{Log}\ (0, 1)$ on $C$ for $\alpha = 1$. Set $f(z) = u(z) + iv(z)$, where $v(z)$ is conjugate[5] to $u(z)$ in $\bar{C}$. Then for $k \geqq 1$, $f(z) \in L_A(k, \alpha)$ on $\bar{C}$ if $0 < \alpha < 1$ and $f(z) \in \text{Log}_A\ (k, 1)$ on $\bar{C}$ if $\alpha = 1$.*

THEOREM 3. *In addition to the hypothesis of Theorem 2 require that the equation of $C$ can be written in the form $z = t(s)$, $s$ arc length measured along $C$, where $t^{(k)}(s)$ exists and satisfies a Lipschitz condition of order $\alpha$. Then if $0 < \alpha < 1$, $u(z) \in L(k, \alpha)$ on $C$; if $\alpha = 1$, $u(z) \in \text{Log}\ (k, 1)$ on $C$.*

To connect the continuity properties of $u(z)$ on $C_R$ and degree of convergence of approximating polynomials on $\bar{C}$ we have

THEOREM 4. *Let $C$ and $C_R$, $R > 1$, be contours.[6] Let $u(z) \in L(k, \alpha)$, $k \geqq 0$, $0 < \alpha \leqq 1$, on $C_R$. Then for each, $n$, $n = 1, 2, \ldots$, there exists a harmonic polynomial $p_n(z)$ such that*

$$|u(z) - p_n(z)| \leqq L/n^{k+\alpha}R^n, \qquad z \text{ on } \bar{C}.$$

THEOREM 5. *Let $C$ be a contour and let $C_R$, $R > 1$, be a Jordan curve. Let $u(z) \in L(k, \alpha)$, $k \geqq 0$, $0 < \alpha \leqq 1$, on $C_R$. Set $f(z) = u(z) + iv(z)$, where $v(z)$ is conjugate to $u(z)$ in $\bar{C}_R$. Then there exist polynomials $\pi_n(z)$ in $z$ such that*

$$|f(z) - \pi_n(z)| \leqq M/n^{k+\alpha}R^n, \qquad z \text{ on } \bar{C}.$$

Let $C$ be a rectifiable Jordan curve lying in the $z$-plane. Let $w = \Omega(z)$ with inverse $z = \chi(w)$ map the interior of $C$ one-to-one and conformally onto the interior of $|w| = 1$. We say that $C$ is of Type $A_1$ if $\chi'(w)$ is continuous and different from zero on $|w| = 1$.

As a converse (not exact) to Theorems 4 and 5 we have

THEOREM 6. *Let $C$ be a curve of Type $A_1$; suppose that $u(z)$ is defined on $\bar{C}$. For each $n$, $n = 1, 2, \ldots$, let a harmonic polynomial $p_n(z)$ exist such that*

$$|u(z) - p_n(z)| \leq M/n^{k+\alpha+1}R^n, \qquad z \text{ on } \bar{C},\ R > 1,\ k \geq 0.$$

*Set $f(z) = u(z) + iv(z)$, where $v(z)$ is conjugate to $u(z)$ in $\bar{C}$. Then $f(z) \in L_A(k, \alpha)$ on $\bar{C}_R$ if $0 < \alpha < 1$ and $f(z) \in \mathrm{Log}_A\ (k, 1)$ on $\bar{C}_R$ if $\alpha = 1$; $u(z) \in L(k, \alpha)$ on $C_R$ if $0 < \alpha < 1$ and $u(z) \in \mathrm{Log}\ (k, 1)$ on $C_R$ if $\alpha = 1$.*

A classification of functions analytic in the unit circle due to Hardy and Littlewood expressed in terms of asymptotic inequalities was used by Walsh and Sewell[7] as the basis of a study of approximation to analytic functions by polynomials in $z$. This same classification of analytic functions yields by studying real and pure imaginary parts a corresponding classification of functions harmonic in the unit circle.

If $u(z)$, $z = re^{i\theta}$, is harmonic in $|z| < 1$ and if

$$|u(re^{i\theta})| \leq L(1 - r)^{k+\alpha}, \qquad r < 1,\ 0 < \alpha \leq 1,$$

where $k < 0$, then we say $u(z) \in L(k, \alpha)$ on $|z| = 1$.

Let $C$ be an analytic Jordan curve in the $z$-plane. Let $w = \Omega(z)$ with inverse $z = \chi(w)$ map the interior of $C$ conformally onto the interior of $|w| = 1$. Then $u(z)$ is said to be of class $L(k, \alpha)$, $k < 0$, on $C$ if $u(\chi(w))$ is of class $L(k, \alpha)$ on $|w| = 1$, $0 < \alpha \leq 1$.

THEOREM 7. *Let $C$ be an analytic Jordan curve. Let $u(z) \in L(k, \alpha)$, $k < 0$, $0 < \alpha \leq 1$, on $C_R$, $R > 1$. Set $f(z) = u(z) + iv(z)$, where $v(z)$ is conjugate to $u(z)$ interior to $C_R$. Then for each $n$, $n = 1, 2, \ldots$, there exists a polynomial $\pi_n(z)$ in $z$ such that*

$$|f(z) - \pi_n(z)| \leq M/n^{k+\alpha}R^n, \qquad z \text{ on } \bar{C}.$$

As a converse (not exact) to Theorem 7 we have

THEOREM 8. *Let $C$ be an analytic Jordan curve, and let $u(z)$ be defined on $\bar{C}$. For each $n$, $n = 1, 2, \ldots$, let a harmonic polynomial $p_n(z)$ exist such that*

$$|u(z) - p_n(z)| \leq M/n^{k+\alpha+1}R^n, \qquad z \text{ on } \bar{C},\ R > 1.$$

*If $k < 0$, $0 < \alpha \leq 1$, $k + \alpha \neq 0$, then $u(z) \in L(k, \alpha)$ on $C_R$.*

For the case $k + \alpha = 0$, a slightly less precise result exists. Theorems 1–8 are in a sense the best possible.

[*] Dr. H. M. Elliott is a Vassie James Hill A.A.U.W. Fellow

[1] By a harmonic polynomial $p_n(z)$ of degree $n$ we mean the real part of a polynomial in $z$ of degree $n$. The degree of a polynomial will be indicated by its subscript.

[2] For these results see Walsh, J. L., *Bull. Am. Math. Soc.*, **35**, 499–544 (1929); Walsh, J. L., *Ann. Math.*, **38**, 321–354 (1937).

[3] A function $u(z)$ is said to satisfy a Lipschitz condition of order $\alpha$ on a set $E$ if for any two points $z_1$, $z_2$ of $E$ we have $|u(z_1) - u(z_2)| \leq L|z_1 - z_2|^{\alpha}$, where $L$ is a constant independent of $z_1$ and $z_2$. The letters $L$ and $M$ henceforth denote constants which may vary from one inequality to another and which may depend on the point sets involved, but which are independent of $z$ and $n$. The letter $k$ will always denote an integer.

[4] Cf. Sewell, W. E., "Degree of Approximation by Polynomials in the Complex Domain," *Annals of Mathematics Studies*, vol. 9, Princeton, 1942.

[5] If two functions are conjugate in a region and continuous in the corresponding closed region, we say that the functions are conjugate in the closed region.

[6] See Sewell, *loc. cit.*, p. 58, for the definition of a contour.

[7] Walsh, J. L., and Sewell, W. E., *Trans. Am. Math. Soc.*, **49**, 229–257 (1941).

---

# *A SIMPLE ANALYTICAL PROOF OF A FUNDAMENTAL PROPERTY OF BIRATIONAL TRANSFORMATIONS*

BY OSCAR ZARISKI

DEPARTMENT OF MATHEMATICS, HARVARD UNIVERSITY

Communicated November 5, 1948

*1. Valuations in Local Domains.*—Let $\mathfrak{o}$ be a local domain, i.e., a local ring which is at the same time an integral domain, and let $\mathfrak{m}$ be the maximal ideal in $\mathfrak{o}$. We consider a valuation $v$ of the quotient field of $\mathfrak{o}$. Let $R$ be the valuation ring of $v$ and let $\mathfrak{P}$ be the maximal ideal of $R$. We say that *$v$ has center $\mathfrak{m}$ in $\mathfrak{o}$* if (1) $\mathfrak{o} \subset R$ and (2) $\mathfrak{m} = \mathfrak{P} \cap \mathfrak{o}$. If $v$ has center $\mathfrak{m}$ in $\mathfrak{o}$, then the residue field $\mathfrak{o}/\mathfrak{m}$ of $\mathfrak{o}$ can be regarded as a subfield of the residue field $R/\mathfrak{P}$ of the valuation $v$. The transcendence degree of the latter field over the former will be called the *$\mathfrak{o}$-dimension* of $v$.

THEOREM 1. *If the local domain $\mathfrak{o}$ is of dimension $r$ and if a valuation $v$ of the quotient field of $\mathfrak{o}$ has center $\mathfrak{m}$ in $\mathfrak{o}$ and has rank 1, then the $\mathfrak{o}$-dimension of $v$ is at most $r - 1$.*

*Proof.*—We shall first consider the case in which $\mathfrak{o}$ is complete. By a theorem due to I. S. Cohen[2] (Theorem 11, p. 79) $\mathfrak{o}$ contains, then, a coefficient ring $\Omega$. In the equal-characteristic case (i.e., if $\mathfrak{o}$ and $\mathfrak{o}/\mathfrak{m}$ have the same characteristic) $\Omega$ is a *field* which is mapped isomorphically mod $\mathfrak{m}$ on the entire residue field $\mathfrak{o}/\mathfrak{m}$. It is this case that occurs in algebraic geometry. In the unequal-characteristic case (i.e., if $\mathfrak{o}$ is of characteristic 0 and $\mathfrak{o}/\mathfrak{m}$ is of characteristic $p \neq 0$) $\Omega$ is a complete discrete valuation ring of rank 1 in which the maximal ideal is generated by the prime number $p$ and which is mapped mod $\mathfrak{m}$ (and therefore also mod $p$) on the entire residue field $\mathfrak{o}/\mathfrak{m}$.

Let $t_1, t_2, \ldots, t_r$ be a system of parameters in $\mathfrak{o}$, i.e., a set of $r$ elements of $\mathfrak{m}$ such that the ideal $\mathfrak{o}\,(t_1, t_2, \ldots, t_r)$ is primary for $\mathfrak{m}$. Then $\mathfrak{o}$ contains the ring $\Omega < t_1, t_2, \ldots, t_r >$ of formal power series in the $t$'s, with coefficients in $\Omega$. We denote this ring by $\Omega^*$. The ring $\Omega^*$ is a complete local domain of dimension $r$, has the same residue field as $\mathfrak{o}$, and $\mathfrak{o}$ is a finite module over $\Omega^*$ (see Chevalley,[1] Lemma 1 and Prop. 5, p. 702. For the general case, see I. S. Cohen,[2] Lemma 16, p. 90). If $v'$ denotes the valuation of the quotient field of $\Omega^*$ induced by $v$ and if $\mathfrak{m}'$ denotes the maximal ideal in $\Omega^*$, then it is clear that $v'$ has center $\mathfrak{m}'$ in $\Omega^*$ and that the residue field of $v$ is an algebraic extension of the residue field of $v'$ (since the quotient field of $\mathfrak{o}$ is an algebraic extension of the quotient field of $\Omega^*$). It follows that it is sufficient to prove the theorem for $\Omega^*$ and $v'$. We therefore assume that $\mathfrak{o}$ itself is a power series ring $\Omega < t_1, t_2, \ldots, t_r >$ over the coefficient ring $\Omega$.

Let $\xi = u/w$ be an arbitrary element in the valuation ring of $v$, where $u$ and $w$ are elements of $\mathfrak{o}$. Let $u_i$ and $w_i$ be the partial sums of the power series $u$ and $w$, respectively. We have $u - u_i \in \mathfrak{m}^i$, $w - w_i \in \mathfrak{m}^i$. Since $v$ has center $\mathfrak{m}$ in $\mathfrak{o}$ and since the rank of $v$ is 1 (i.e., the value group of $v$ consists of real numbers), it follows that $v(u) = v(u_i)$ and $v(w) = v(w_i)$ for all $i$ sufficiently large. We have $\xi - u_i/w_i = [w_i(u - u_i) - u_i(w - w_i)]/ww_i$. As $i$ becomes very large the value of the numerator (in the given valuation $v$) approaches $+\infty$, while the value of the denominator remains constantly equal to the value of $w^2$. Hence we have, for $i$ very large, $v(\xi - u_i/w_i) > 0$, and therefore $\xi$ and $u_i/w_i$ have the same residue in the given valuation $v$. Since $u_i$ and $w_i$ are polynomials in $\Omega[t_1, t_2, \ldots, t_r]$, it follows that the residue field of $v$ is the same as the residue field of the valuation $v_1$ induced by $v$ in the ring $S$ of rational functions $\Omega(t_1, t_2, \ldots, t_r)$, and from this our theorem follows at once. For in the equal-characteristic case $\Omega$ and $s$ are fields, $v_1$ is trivial on $\Omega$ and is non-trivial on $S$. In the unequal-characteristic case we may assume that $t_1 = p$. In that case $S$ has transcendence degree $r - 1$ over the quotient field of $\Omega$, and it is clear that the transcendence degree, over $\Omega/\Omega\,p$, of the residue field of $v_1$ cannot exceed $r - 1$.

We now can easily prove the theorem for arbitrary (not necessarily complete) local domains $\mathfrak{o}$. Let $\mathfrak{o}^*$ and $R^*$ be the $\mathfrak{m}$-adic and the $\mathfrak{P}$-adic completions of $\mathfrak{o}$ and of the valuation ring $R$, respectively. Since $v$ is of rank 1, the completion of $R$, in the sense of the theory of local rings, coincides with the completion of $R$, in the sense of valuation theory. Hence, $R^*$ is a complete valuation ring for an extension $v^*$ of $v$, and $v^*$ has the same value group and the same residue field as $v$. The assumption that $v$ has center $\mathfrak{m}$ in $\mathfrak{o}$ implies that Cauchy sequences and zero sequences in $\mathfrak{o}$ are, respectively, also Cauchy sequences and zero sequences in $R$. We have therefore a natural homomorphism $\tau$ of $\mathfrak{o}^*$ into $R^*$. Let $\mathfrak{o}' = \tau\mathfrak{o}^*$.

Then $\mathfrak{o}'$ is a complete local domain, and we have $\mathfrak{o} \subset \mathfrak{o}'$, since $\tau$ reduces to the identity on $\mathfrak{o}$. If $\mathfrak{m}^*$ and $\mathfrak{m}'$ are the maximal ideals in $\mathfrak{o}^*$ and $\mathfrak{o}'$, respectively, then $\mathfrak{m}' = \tau\mathfrak{m}^*$, $\mathfrak{o}'/\mathfrak{m}' = \mathfrak{o}^*/\mathfrak{m}^*$, and since $\mathfrak{m}^* = \mathfrak{o}^*\mathfrak{m}$ and $\mathfrak{o}^*/\mathfrak{m}^* = \mathfrak{o}/\mathfrak{m}$, we conclude that $\mathfrak{m}' = \mathfrak{o}'\cdot\mathfrak{m}$, $\mathfrak{m}' \cap \mathfrak{o} = \mathfrak{m}$ and $\mathfrak{o}'/\mathfrak{m}' = \mathfrak{o}/\mathfrak{m}$. Finally we point out that $\mathfrak{o}'$ is of dimension $\leqq r$, since $\mathfrak{o}^*$ is exactly of dimension $r$. Now, let $v'$ be the valuation of the quotient field of $\mathfrak{o}'$ induced by $v^*$. Since $\mathfrak{o}' \subset R^*$ and $\mathfrak{m}' = \mathfrak{o}'\mathfrak{m}$, it follows that the valuation $v'$ (which is of rank 1) has center $\mathfrak{m}'$ in $\mathfrak{o}'$. Therefore by the previous case, the $\mathfrak{o}'$-dimension of $v'$ is at most $r - 1$. Since $v'$ is an extension of $v$ and since the residue fields of $\mathfrak{o}$ and $\mathfrak{o}'$ are the same, it follows *a fortiori* that the $\mathfrak{o}$-dimension of $v$ is at most $r - 1$. This establishes the theorem for arbitrary local domains.

The second part of the proof yields an important consequence in the following special case: *the local domain $\mathfrak{o}$ is analytically irreducible, i.e., $\mathfrak{o}^*$ is an integral domain, and the valuation $v$ (of rank 1 and center $\mathfrak{m}$) is of dimension $r - 1$.* In that case, also, $v'$ must be exactly of dimension $r - 1$. In view of Theorem 1, this is possible only if in the inequality $\dim \mathfrak{o}' \leqq r$ the equality sign holds. Hence $\mathfrak{o}'$ is of dimension $r$. Since also the local *domain* $\mathfrak{o}^*$ is of dimension $r$, *it follows that the homomorphism* $\tau$ of $\mathfrak{o}^*$ *onto* $\mathfrak{o}'$ *must be an isomorphism.* We can therefore state the following theorem:

THEOREM 2. *Let $\mathfrak{o}$ be a local domain of dimension $r$ and let $\mathfrak{m}$ be the maximal ideal of $\mathfrak{o}$. Let $v$ be a rank 1 valuation of the quotient field of $\mathfrak{o}$ such that the $\mathfrak{o}$-dimension of $v$ is $r - 1$ and such that $\mathfrak{m}$ is the center of $v$ in $\mathfrak{o}$. If $\mathfrak{o}$ is analytically irreducible (i.e., if the completion $\mathfrak{o}^*$ of $\mathfrak{o}$ is an integral domain), then $\mathfrak{o}$ is a subspace of the valuation ring $R$ of $v$, i.e., we have: $\mathfrak{P}^i \cap \mathfrak{o} \subset \mathfrak{m}^{\nu(i)}$, $\nu(i) \to +\infty$, as $i \to +\infty$ where $\mathfrak{P}$ is the maximal ideal of $R$. The valuation $v$ can then be extended to a valuation of $\mathfrak{o}^*$.*

2. *An Application to Algebraic Varieties.*—Let $V$ and $V'$ be two algebraic varieties over a common ground field $k$, and assume that the function field $\Sigma$ of $V$ is contained in the function field $\Sigma'$ of $V'$ (whence $V$ is a rational transform of $V'$). Let $P$ and $P'$ be two corresponding points of $V$ and $V'$.

THEOREM 3. *If the local ring of $P'$ contains the local ring of $P$ and if $V$ is locally irreducible at $P$, then the local ring of $P$ is a subspace of the local ring of $P'$.*

*Proof:* Let $\mathfrak{o}$ and $\mathfrak{o}'$ denote the local rings of $P$ and $P'$, respectively. Let $\mathfrak{m}$ and $\mathfrak{m}'$ be the maximal ideals in these two rings. We have, by assumption, $\mathfrak{o} \subset \mathfrak{o}'$. Furthermore, the fact that $P$ and $P'$ are corresponding points signifies that $\mathfrak{m} \subset \mathfrak{m}'$ (and hence $\mathfrak{m}' \cap \mathfrak{o} = \mathfrak{m}$). Let $r$ and $r'$ be the dimensions of $\mathfrak{o}$ and $\mathfrak{o}'$, respectively. Let $s = \dim V$, $s' = \dim V'$. Note that the difference $s - r$ is merely the dimension of the point $P$ over $k$, i.e., the dimension of the irreducible variety of which $P$ is a general point. Similarly for $s'$ and $r'$. We fix a divisor $v'$ [i.e., an $(s' -$

1) dimensional valuation $v'$] of the function field $\Sigma'$ which has center at the point $P'$. The induced valuation $v$ in the function field $\Sigma$ has then center $P$, and its dimension (the dimensions of $v$ and $v'$ are intended here with reference to $k$) is necessarily $s - 1$ (since dim $v'$ − dim $v$ cannot exceed $s' - s$). In the terminology of the preceding section, $v$ has center $\mathfrak{m}$ in $\mathfrak{o}$, and its $\mathfrak{o}$-dimension is equal to $r - 1$. Similarly $v'$ has center $\mathfrak{m}'$ in $\mathfrak{o}'$, and its $\mathfrak{o}'$-dimension is equal to $r' - 1$. Let $\mathfrak{P}$ and $\mathfrak{P}'$ be the prime ideals of the valuations $v$ and $v'$. It is clear that high powers of $\mathfrak{P}'$ contract in the valuation ring of $v$ to high powers of $\mathfrak{P}$. Since $V$ is analytically irreducible at $P$, it follows from Theorem 2 that high powers of $\mathfrak{P}$ contract in $\mathfrak{o}$ to high powers of $\mathfrak{m}$. Therefore high powers of $\mathfrak{P}'$ contract in $\mathfrak{o}$ to high powers of $\mathfrak{m}$. But since $\mathfrak{m}' \subset \mathfrak{P}'$, it follows *a fortiori* that high powers of $\mathfrak{m}'$ contract in $\mathfrak{o}$ to high powers of $\mathfrak{m}$, i.e., the local ring $\mathfrak{o}$ is a subspace of the local ring $\mathfrak{o}'$, q. e. d.

From Theorem 3 it is possible to derive in a very simple manner the "main theorem" of our paper on birational transformations (Zariski,[3] p. 522). Let $V$ and $V'$ be birationally equivalent varieties and let $T$ denote the birational transformation of $V$ onto $V'$. We denote by $T\{P\}$ the set of points of $V'$ which correspond to $P$. We say that a point $P'$ of $T\{P\}$ is *isolated* if $P'$ is not a specialization (over $k$) of any point $Q'$ of $T\{P\}$ such that dim $Q'/k >$ dim $P'/k$.

THEOREM 4. *Under the assumptions of Theorem 3 and under the additional assumptions that $P'$ is an isolated point of $T\{P\}$ and that dim $P/k$ = dim $P'/k$, the set $T\{P\}$ consists of a finite number of points, any two of which are isomorphic over $k$.*

*Proof:* We use the notation of the proof of Theorem 3. Let $\mathfrak{o}^*$ and $\mathfrak{o}'^*$ be the completions of $\mathfrak{o}$ and $\mathfrak{o}'$, respectively. Since, by Theorem 3, $\mathfrak{o}$ is a subspace of $\mathfrak{o}'$, it follows that $\mathfrak{o}^*$ is also a subring and subspace of $\mathfrak{o}'^*$. The assumption that $P'$ is an isolated point of $T\{P\}$ signifies that the ideal $\mathfrak{o}' \cdot \mathfrak{m}$ is primary for $\mathfrak{m}'$. Since, on the other hand, $\mathfrak{o}$ and $\mathfrak{o}'$ have, by assumption, the same dimension, it follows that if $\{t_1, t_2, \ldots, t_r\}$ is a system of parameters in $\mathfrak{o}$, then the $t$'s form also a system of parameters in $\mathfrak{o}'$. Therefore $\mathfrak{o}'^*$ is a finite module over the power series ring $k < t_1, t_2, \ldots, t_r >$ (Chevalley[1]), and *a fortiori* $\mathfrak{o}'^*$ is a finite module over $\mathfrak{o}^*$. In particular, *every element of $\mathfrak{o}'$ is integrally dependent on $\mathfrak{o}^*$.*

Now let $v$ be any divisor of the function field $\Sigma$, with center at $P$. By Theorem 2, $v$ can be extended to a valuation $v^*$ of the quotient field of the local *domain* $\mathfrak{o}^*$. Since every element of $\mathfrak{o}'$ is integrally dependent on $\mathfrak{o}^*$, it follows that $\mathfrak{o}'$ is contained in the valuation ring of $v^*$ (since $\mathfrak{o}^*$ is contained in that valuation ring). Consequently $\mathfrak{o}'$ *is contained in the valuation ring of $v$* (since $v^*$ is an extension of $v$). The center of the divisor $v$ on the variety $V'$ is therefore a point $A'$ such that *the local ring of $A'$ contains the local ring* $\mathfrak{o}'$ of $P'$, i.e., a point $A'$ such that $P'$ is a specialization of $A'$ over $k$. Since $P$ and $A'$ are centers of $v$, $A'$ belongs to the set $T\{P\}$. Since $P'$ is an isolated point of $T\{P\}$, we must have $\dim A'/k = \dim P'/k$. From this it follows that $A'$ and $P'$ are isomorphic points over $k$ (since $P'$ is a specialization of $A'$ over $k$) and that consequently the local rings of $A'$ and $P'$ coincide. Since $\mathfrak{o} \subset \mathfrak{o}'$, the local field $k(A')$ contains the local field $k(P)$ of $P$, and hence $A'$ is an algebraic point over $k(P)$ (since $\dim A'/k = \dim P'/k = \dim P/k$). Since $T\{P\}$ is a variety over $k(P)$, we conclude that $T\{P\}$ is zero dimensional over $k(P)$ and therefore consists of a finite number of points. This completes the proof.

Our "Main Theorem" on birational transformations follows from the above theorem if one takes into account that if $V$ is locally normal at $P$ then (*a*) $V$ is analytically irreducible at $P$ (Zariski[4]) and (*b*) the *finite* set $T\{P\}$ necessarily consists of a single point (Zariski[3], theorem 8(A) and theorem 10, pp. 512–514).

[1] Chevalley, C., "On the Theory of Local Rings," *Ann. Math.*, **44**, 690–708 (1943).

[2] Cohen, I. S., "On the Structure and Ideal Theory of Complete Local Rings," *Trans. Am. Math. Soc.*, **59**, 54–106 (1946).

[3] Zariski, O., "Foundations of a General Theory of Birational Correspondences," *Ibid.*, **53**, 490–542 (1943).

[4] Zariski, O., "Analytical Irreducibility of Normal Varieties," *Ann. Math.*, **49**, 352–361 (1948).

---

# THE NUCLEOPROTEIN CONTENT OF WHOLE NUCLEI

By A. W. Pollister and C. Leuchtenberger

Department of Zoology, Columbia University

Communicated by L. C. Dunn, December 3, 1948

The task of determining the chemical composition of parts of a cell can be approached in two different ways, which may for convenience be called the *gross* and the *cytological* methods. The former is followed by the analytical biochemist, who isolates cellular structures in masses (millions or billions of units) and subjects the mass to a variety of analytical procedures. The cytologist, by contrast, studies with his microscope single cells, and for chemical analysis of these structures he is almost wholly restricted to methods which are based upon light absorption. Obviously the gross and cytological methods are complementary in many respects, as, for example, in evaluating the changes which may have been brought about by the isolation procedure. The nucleoprotein content of the cell nucleus has been determined in recent years by both gross and cytological methods. In the present paper we are concerned with demonstrating and attempting to explain the considerable differences between the results from these two lines of investigation.

Gross analyses of nucleic acid and protein in isolated nuclei and chromosomes have been made in recent years by Claude and Potter, by Mirsky and his associates and by the Stedmans. Although in methods of isolation and analysis these three groups have worked very differently, their results are in essential agreement on the over-all nucleoprotein composition, as shown by the ratio of protein to nucleic acid, table 1. In all these interphase nuclei and chromosomes gross analyses showed the protein content to be not more than one or two times that of the nucleic acid.

We have determined the protein-nucleic acid ratio in two of these nuclear types by microscopic methods applied to cytological sections of fixed material. For desoxypentose nucleic acid, the Feulgen reaction was estimated from measurement of the absorption of light isolated from a tungsten lamp by a Farrand Interference Filter with peak transmission at 550 mμ. Nucleic acid was computed from these values (called $E_{550}$ in the tables) by the method described by Di Stefano.[4] For protein estimation the Millon reaction was carried out and measured by the total protein method of Pollister and Ris.[5] From these extinction values for the tyrosine-plus-tryptophane mercurials (called $E_{365}$ in the tables) protein was computed by assuming that the absorbing substance was all tyrosine, and that the tyrosine constituted one-sixteenth (6.25%) of the mass of the protein. In nuclei the absorbing material is not homogeneously distributed. Absorption measurements of such structures always involve an error, not easily estimable, resulting in too low extinction values for the absolute amount of material in the absorption path. However, this has little effect upon the ratio between the amounts of two substances with identical distributions, a condition we have assumed for protein and nucleic acid in these nuclei.

TABLE 1

TOTAL PROTEIN AND NUCLEIC ACID IN CHROMOSOMES AND NUCLEI WHICH HAVE BEEN SUBJECTED TO ISOLATION PROCEDURES. THE RATIOS IN SOME CASES HAVE BEEN COMPUTED FROM GROSS ANALYTICAL DATA IN THE REFERENCES

| MATERIAL | REFERENCE | PROTEIN / NUCLEIC ACID |
|---|---|---|
| Chromosomes, thymus lymphocytes (calf) | Mirsky and Ris[1] | 1.5 |
| Nuclei, thymus lymphocytes (human) | Stedman and Stedman[2] | 1.35 |
| Nuclei, liver (human) | Stedman and Stedman[2] | 2.6 |
| Nuclei, liver (ox) | Stedman and Stedman[2] | 2.3 |
| Nuclei, liver (mouse) | Stedman and Stedman[2] | 2.4 |
| Chromosomes, leukemic cells (mouse) | Claude and Potter[3] | 1.5 |

From these cytological estimates, table 2, it appears that in the nuclei of fixed tissues the protein is very high in proportion to the nucleic acid, the latter forming only about one-ninth of the nucleoprotein mass of the thymus nucleus and one-eighteenth of the liver nucleus. A similar composition of the interphase nucleus is indicated by cytological determinations which have been made on various nuclei, by us and other workers, table 3. In short, all microscopical determinations of nucleic acid and protein agree in demonstrating that at interphase, or early prophase, the nucleus contains a very large amount of protein in proportion to the nucleic acid. This is in sharp disagreement with the equally uniform results of gross analyses, by different workers, of chromosomes and nuclei which have been subjected to a procedure for isolating them from the rest of the cell. In all of the latter a very much lower proportion of protein was found.

TABLE 2

TOTAL PROTEIN AND NUCLEIC ACID IN NUCLEI OF FIXED SECTIONS, DETERMINED BY CYTOLOGICAL ABSORPTION MEASUREMENTS

| MATERIAL | FEULGEN REACTION $E_{546}$ | MILLON REACTION $E_{365}$ | PROTEIN / NUCLEIC ACID |
|---|---|---|---|
| Thymus (guinea pig) | 0.343 ± 0.04 | 0.086 ± 0.005 | 8.4 |
| Liver (mouse) | 0.413 ± 0.02 | 0.209 ± 0.02 | 16.9 |

TABLE 3

PROTEIN-NUCLEIC ACID RATIO IN VARIOUS NUCLEI, FOUND BY CYTOLOGICAL ABSORPTION MEASUREMENTS

| MATERIAL | REFERENCE | METHODS<br>NUCLEIC ACID | <br>PROTEIN | PROTEIN / NUCLEIC ACID |
|---|---|---|---|---|
| *Lophius* nerve cell, nucleus | Hyden[6] | Ultraviolet absorption, $E_{2537}$ | Ultraviolet absorption, $E_{2800}$ | 14 |
| Amblystoma cartilage cell, nucleus | Di Stefano[4] | Ultraviolet absorption, $E_{2537}$ | Millon reaction, $E_{365}$ | 10 |
| *Drosophila* salivary gland chromosome | Caspersson[7] | Ultraviolet absorption, $E_{2537}$ | Ultraviolet absorption, $E_{2800}$ | 10 |
| Grasshopper spermatocyte, early meiotic prophase nucleus | Caspersson[8] | Ultraviolet absorption, $E_{2537}$ | Ultraviolet absorption, $E_{2800}$ | 20 |
| (*Chortophaga*) grasshopper spermatocyte, leptotene nucleus | Ris[9] | Ultraviolet absorption, $E_{2537}$ | Millon reaction, $E_{365}$ | 11* |
| (*Chortophaga*) grasshopper spermatocyte, leptotene nucleus | Pollister and Leuchtenberger | Feulgen reaction, $E_{580}$ | Millon reaction, $E_{365}$ | 10 |
| Onion root tip, meristem, nucleus | Ris[9] | Ultraviolet absorption, $E_{2537}$ | Millon reaction, $E_{365}$ | 6.5* |

* Computed by assuming the $E_{365}$ represents absorption due to tyrosine, which is 6.25% of the protein.

One can reconcile these divergent analyses by assuming that the isolation process has removed from the nuclei a large proportion of the protein. An essential part of all the methods of getting a mass of "pure nuclei" for chemical analysis is a prolonged washing, with such reagents as citric acid or physiological saline. The latter has been demonstrated to wash out a large amount of protein from the cytoplasm (Mirsky and Pollister[10]), and it is logical to suppose that loss of protein from the nuclei also may result from this sort of treatment. This point was investigated on the liver of the guinea pig. Slices of fresh liver were cut on a freezing microtome set at 100 microns, some were fixed immediately for controls, others were kept in physiological saline at 5°C. before fixation. Protein was determined by photometric measurement of the Millon reaction on paraffin sections of the fixed tissue slices. These measurements showed that in a slice that had been in normal saline for three hours the nuclei contained on the average about one-fourth as much protein as the nuclei of the untreated control slice. This same saline treatment reduced the cytoplasmic protein content to nearly one-half its original value. The suspension of a thick slice of liver in saline is a mild physical and chemical treatment in comparison with, for example, the homogenization in a Waring Blendor and subsequent washing, which are part of one of the procedures followed in isolating nuclei or chromosomes for gross analysis (Mirsky and Ris[1]). There seems good reason to suppose that the more drastic procedures would reduce the protein content of nuclei and chromosomes to the values given in table 1.

The cytological evidence shows that in nuclei of sections of fixed tissues in which rapid cytological fixation has preserved the nuclei with little chance for loss of protein or nucleic acid, there is $6^1/_2$ to 20 times as much protein as nucleic acid, and furthermore suggests that the process of isolation for gross analysis washes out most of this protein from the nuclei. Without more cytological or chemical data it cannot be said whether the protein removed by washing was originally a part of the chromosomes or some component of the interchromosomal material. There is clear indication that the excess protein in whole (i.e., fixed) nuclei is not histone. Mirsky and Ris[1], have reported that histone constitutes over 80% of the protein of isolated thymus chromosomes, and this ratio was approximately verified by microscopic absorption measurements of isolated nuclei (Pollister and Ris[5]). If histone likewise constitutes four-fifths of the protein of whole nuclei, its removal should lead to a corresponding reduction in the Millon reaction. Instead, however, when sections of fixed thymus are treated by the reagent which dissolves histone (sulphuric acid plus divalent mercury) and the Millon reaction ($E_{365}$) on these sections is compared with that of nuclei in untreated tissue, then it is found that the decrease of $E_{365}$ caused by histone removal is barely significant, table 4. This is actually about the result to be expected if the histone content of whole thymus nuclei is that reported by Mirsky and Ris for isolated chromosomes (1.25 times the nucleic acid, which would be 15% of the protein of whole nuclei, and equivalent to an $E_{365}$ of 0.021). These data show that in the large amount of protein of whole nuclei there is probably very little more histone than has been found in the same type of nucleus after isolation. In other words, the isolation procedure appears to cause little if any loss of histone, There is some evidence that the desoxypentose nucleic acid content likewise is not changed appreciably during isolation of chromosomes.

TABLE 4

EFFECT OF REMOVAL OF HISTONE ON THE PROTEIN CONTENT OF THYMUS NUCLEI AS SHOWN BY THE MILLON REACTION. (THE VALUES ARE FROM MEASUREMENTS THROUGH THE CENTRAL PART OF THE NUCLEUS, WITH A CONSEQUENT LONGER OPTICAL PATH AND HIGHER EXTINCTION THAN IN TABLE 2)

| TOTAL PROTEIN | | NON-HISTONE PROTEIN (AFTER $H_2SO_4$ + $Hg^{++}$) | |
|---|---|---|---|
| NUMBER OF NUCLEI | MEAN $E_{365}$ | NUMBER OF NUCLEI | MEAN $E_{365}$ |
| 31 | 0.137 ± 0.004 | 30 | 0.123 ± 0.005 |

We can only speculate as to the other characteristics of this non-histone excess protein of whole nuclei. Perhaps it should be classed with the other non-histone protein fractions of the nucleus, and the "chromosomin" of the Stedmans and the "residual protein" of Mirsky and Ris be regarded as residues, which for some reason do not go into solution when nuclei are isolated for gross analysis. Since the metabolically inactive spermatozoan nucleus appears to be almost entirely nucleoprotamine, or nucleohistone (Pollister and Mirsky[11]), it seems logical to consider the non-histone, which is so abundant in metabolically active cells, to be a part of the nucleus which is concerned with the special metabolic functions of the nucleus, leaving to the nucleohistone the general function of chromosome reproduction. Thus, in the non-histone protein fraction must be included the enzymes which were found by Dounce.[12] Indeed, Dounce showed that certain steps of the nuclear isolation process do remove significant amounts of protein and enzymes from the nuclei. Finally, it is intriguing also to hope that in the excess protein fraction may be found some of the elusive intermediate products of protein synthesis, which Caspersson has suggested may be the primary intranuclear metabolism.

The results may be briefly summarized. Photometric measurements of nuclei in sections of fixed tissue show that these contain a much higher proportion of protein than has been reported in nuclei and chromosomes when isolated for gross chemical analysis. This high protein content is consistent with other cytological estimates of nucleoprotein in fixed nuclei. It is shown that a large amount of protein is removed from nuclei by treatment with physiological saline, which is a part of one method of isolating chromosomes. The protein which occurs in whole nuclei and which is apparently lost during the isolation procedure does not appear to include any histone.

[1] Mirsky, A. E., and Ris, H., *J. Gen. Phys.*, **31**, 7–18 (1947).

[2] Stedman, E., and Stedman, E., *Cold Spring Harbor Symposia*, XII, 224–236 (1947).

[3] Claude, A., and Potter, J. S., *J. Exp. Med.*, **77**, 345–354 (1943).

[4] Di Stefano, H. S., *Chromosoma*, in press (1948).

[5] Pollister, A. W., and Ris, H., *Cold Spring Harbor Symposia*, XII, 147–157 (1947).

[6] Hyden, H., *Ztschft. mikr. anatom. Forsch.*, **54**, 96–130 (1943).

[7] Caspersson, T., *Chromosoma*, 1, 147–156 (1940).

[8] Caspersson, T., *Salman Lectures* (1948).

[9] Ris, H., *Cold Spring Harbor Symposia*, XII, 158–160 (1947).

[10] Mirsky, A. E., and Pollister, A. W., PROC. NAT. ACAD. SCI., **28**, 344–352 (1942).

[11] Pollister, A. W., and Mirsky, A. E., *J. Gen. Phys.*, **30**, 101–116 (1946).

[12] Dounce, A. L., *J. Biol. Chem.*, **147**, 685–698 (1943).

PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES
Volume 35 February 15, 1949 Number 2

# *EFFECT OF VISIBLE LIGHT ON THE RECOVERY OF STREPTOMYCES GRISEUS CONIDIA FROM ULTRA-VIOLET IRRADIATION INJURY**

By Albert Kelner

The Biological Laboratory, Cold Spring Harbor, New York

Communicated by M. Demerec, December 5, 1948

It is well known that cells rendered non-viable by ultra-violet or x-irradiation[1, 2, 3] may at times regain their viability if stored under suitable conditions after irradiation. In the case of microorganisms the criterion for viability is usually the ability to form a colony on a solid medium. By recovery is meant the restoration of the ability of an irradiated microorganism to grow and form a colony.

Little is known about the mechanism of the recovery phenomenon; experimental results reported in the literature have been extremely variable. Moreover, at best only a small percentage of the cells rendered non-viable in an irradiated population recover their viability—that is, the over-all recovery is usually relatively slight.

During a study of antibiotically active mutants in actinomycetes[4] we observed that the per cent survival of ultra-violet irradiated *Streptomyces griseus* ATC3326 (a non-streptomycin producer) conidia increased about 10-fold when irradiated suspensions were stored one or two days following irradiation. So little was known about the recovery phenomenon, with which our observation was obviously connected, and the implications of this phenomenon to genetics, medicine, and cellular physiology seemed so important to us, that an intensive study of recovery from irradiation was initiated.

Since observers have found recovery to take place when irradiated cells are stored in the cold,[3] and since our own first observations were made on suspensions which had been stored in the ice box, the first study was one on effect of temperature. It was soon clear that recovery was not dependent on storage in the cold. However results were extremely variable even in duplicate experiments; for example, one suspension of ultra-violet irradiated spores showed no recovery upon storage at 35°C., while another suspension prepared from the same lot of spores and irradiated in exactly the same way, showed a 100,000-fold recovery. Some variable factor seemed present in our experiments which overshadowed in importance the effect of temperature *per se* on recovery. Careful consideration was made of variable factors which might have accounted for such tremendous variation. We were using a glass-fronted water bath placed on a table near a window, in which were suspended transparent bottles containing the irradiated spores. The fact that some of the bottles were more directly exposed to light than others suggested that light might be a factor. Moreover, the greatest and most consistent recovery in our preliminary experiments had taken place in suspensions stored in transparent bottles at room temperature on an open shelf exposed to diffuse light from a window. Experiment showed that exposure of ultra-violet irradiated suspensions to light resulted in an increase in survival rate or a recovery of 100,000- to 400,000-fold. Controls kept in the dark (experiments were made between 15°C. to 37°C. only) showed no recovery at all.

The magnitude of the light effect can hardly be overemphasized. The recovery was so much more complete than any previously observed, that we felt we were dealing here with a key factor in the mechanism causing inactivation and recovery from ultra-violet irradiation.

*Methods.*—The ultra-violet source was a General Electric 15-watt germicidal lamp, 80 per cent of whose ultra-violet radiation was at 2537 Å. The spores of *S. griseus* ATC3326 were suspended in saline or distilled water and irradiated in a thin layer in an open petri dish placed under the ultra-violet source. The suspension was shaken gently during irradiation. Preparation of spores, irradiation, and assay for viable count were otherwise similar to those described previously.[4, 5] Following ultra-violet irradiation, the spore suspensions were placed in glass bottles or test tubes and suspended in a thermostatically controlled glass-fronted water bath. Visible light illumination from various sources as described under individual experiments was directed against the suspension The light passed through two glass thicknesses, and about $^1/_2$ cm. of water, before reaching the ultra-violet irradiated cells. Filters were used in later experiments as described below. Counts were made of the viable cells in a suspension by plating on nutrient agar and incubating 3 days at 28°C. Ultra-violet treated cells which were to be kept in the dark were placed in a covered can suspended in the water bath.

*Effect of Dosage of Ultra-Violet Light on Recovery.*—Conidial suspensions were irradiated with ultra-violet at 60 cm. distance from the lamp (intensity about 960 ergs $\times$ min. $^{-1}$ $\times$ mm. $^{-2}$) for periods indicated in table 1. Immediately after irradiation the suspension was divided into two portions, one of which was kept as a dark control, and the other exposed to light from a window about 2 feet away. In this early experiment the visible light source was not controlled, the suspensions being in the dark at night, and subject to variation in light intensity during the day. However this experiment shows well the consistent recovery which occurred only in the light. Non-ultra-violet treated controls were little affected by visible light, there being if anything a decrease in the count in the light-exposed tubes. In no case did the tubes of ultra-violet treated cells kept in the dark show significant recovery, while in all cases the light-exposed tubes showed recovery varying from 14- to over 72,000-fold according to ultra-violet dosage in this experiment. If the decrease in count of the non-ultra-violet irradiated suspension exposed to light is taken into account it is seen that in the 4-minute experiment the recovery is complete by the fifth day, the count in the ultra-violet irradiated suspension (1.8 × $10^5$) equaling that of the non-ultra-violet irradiated suspension (1.7 × $10^5$).

*Effect of Intensity and Duration of Visible Light*

TABLE 1

Effect of Dose of Ultra-Violet Irradiation on Recovery in the Visible Light. Number of Viable Cells per Ml. of Suspension after Holding Irradiated Cells in Light (*L*) and in Dark (*D*)

| ULTRAVIOLET IRRADIATION, MIN. AND SUBSEQUENT TREATMENT | TIME AT 35° SUBSEQUENT TO ULTRA-VIOLET IRRADIATION | | | | | | | MAXIMUM INCREASE IN SURVIVAL RATE |
|---|---|---|---|---|---|---|---|---|
| | 0 | 2 HRS. | 1 DAY | 2 DAYS | 3 DAYS | 4 DAYS | 5 DAYS | |
| 0 *D* | 2.4 × $10^6$ | . . . | 1.5 × $10^6$ | 1.1 × $10^6$ | 6.3 × $10^5$ | 6.3 × $10^5$ | 3.7 × $10^5$ | . . . |
| *L* | 2.4 × $10^6$ | . . . | 1.5 × $10^6$ | 9.1 × $10^5$ | 5.1 × $10^5$ | 2.9 × $10^5$ | 1.7 × $10^5$ | . . . |
| 4 *D* | 6.2 × $10^4$ | . . . | 2.2 × $10^4$ | 1.3 × $10^4$ | 1.1 × $10^4$ | 7.4 × $10^3$ | 9.7 × $10^3$ | . . . |
| *L* | 6.2 × $10^4$ | . . . | 8.6 × $10^5$ | 4.8 × $10^5$ | 2.4 × $10^5$ | 2.0 × $10^5$ | 1.8 × $10^5$ | 14-fold |
| 5 *D* | 5.1 × $10^3$ | . . . | 4.9 × $10^3$ | 4.7 × $10^3$ | 4.4 × $10^3$ | 4.6 × $10^3$ | 4.9 × $10^3$ | . . . |
| *L* | 5.1 × $10^3$ | . . . | 7.4 × $10^5$ | 4.3 × $10^5$ | 1.7 × $10^5$ | 9.1 × $10^4$ | 5.4 × $10^4$ | 145-fold |
| 6 *D* | 60 | . . . | 4.1 × $10^2$ | 15 | 20 | 40 | 55 | . . . |
| *L* | 60 | . . . | 4.9 × $10^5$ | 2.5 × $10^5$ | 2.0 × $10^5$ | 1.6 × $10^5$ | 1.2 × $10^5$ | 8,200-fold |
| 7 *D* | 15 | . . . | 27 | 15 | 3 | 18 | 20 | . . . |
| *L* | 15 | 4.0 × $10^4$ | 1.5 × $10^5$ | 3.7 × $10^4$ | 1.9 × $10^4$ | 1.5 × $10^4$ | 9.3 × $10^3$ | 10,000-fold |
| 8 *D* | <2.5 | . . . | 10 | 20 | 18 | 45 | 40 | . . . |
| *L* | <2.5 | . . . | 1.6 × $10^5$ | 7.4 × $10^4$ | 3.9 × $10^4$ | 2.7 × $10^4$ | 1.9 × $10^4$ | >64,000-fold |
| 9 *D* | <2.5 | . . . | <2.5 | <2.5 | <2.5 | <2.5 | <2.5 | . . . |
| *L* | <2.5 | . . . | 1.8 × $10^5$ | 3.6 × $10^4$ | 2.9 × $10^4$ | 5.8 × $10^4$ | 1.6 × $10^4$ | >72,000-fold |

*Illumination.*—A conidial suspension was irradiated with ultra-violet for $1^1/_2$ minutes at 20 cm. distance from the mercury lamp. Immediately after irradiation it was placed in a 28°C. water bath and exposed to as high an intensity of artificial light as was conveniently possible to obtain in our laboratory (two photoflood lamps and light from a projection lantern, all placed about 30 cm. from the cells). Table 2 shows the extent of recovery after various time periods. The temperature of the cell suspension did not rise more than 2 degrees during the illumination. Recovery is proportional to duration of illumination, within limits.

TABLE 2

EFFECT OF DURATION OF VISIBLE LIGHT ILLUMINATION ON RECOVERY

| ILLUMINATION TIME, MIN. | VIABLE CELLS PER ML. OF SUSPENSION | RELATIVE INCREASE IN SURVIVAL RATE |
|---|---|---|
| 0 | 2.5* | ... |
| 10 | $2.5 \times 10^3$ | 1,000-fold |
| 20 | $9.2 \times 10^3$ | 3,700-fold |
| 30 | $1.3 \times 10^5$ | 52,000-fold |
| 40 | $1.6 \times 10^5$ | 64,000-fold |
| 50 | $2.0 \times 10^5$ | 80,000-fold |
| 60 | $5.3 \times 10^5$ | 210,000-fold |
| 145 | $5.5 \times 10^5$ | 220,000-fold |
| 173 | $7.7 \times 10^5$ | 310,000-fold |
| 240 | $8.0 \times 10^5$ | 320,000-fold |

* The count of the non-ultra-violet irradiated suspension was $4.2 \times 10,^6$ so that the survival rate at time zero was $6.0 \times 10^{-7}$.

TABLE 3

EFFECT OF TEMPERATURE ON RATE OF RECOVERY. ILLUMINATION PERIOD CONSTANT

| | VIABLE CELLS PER ML. AT VARIOUS TEMPERATURES | | | | |
|---|---|---|---|---|---|
| | 20°C. | 25°C. | 30°C. | 35°C. | 40°C. |
| Exp. 1* | $9.6 \times 10^3$ | $3.9 \times 10^4$ | $3.6 \times 10^4$ | $1.0 \times 10^5$ | $1.1 \times 10^5$ |
| Exp. 2† | ... | ... | ... | ... | $2.3 \times 10^5$ |
| | 45°C. | 50°C. | 55°C. | 60°C. | |
| Exp. 1* | ... | ... | ... | ... | |
| Exp. 2† | $2.4 \times 10^5$ | $3.3 \times 10^5$ | $2.9 \times 10^5$ | $2.2 \times 10^5$ | |

* Exp. 1: Count of non-ultra-violet irradiated control was $8.0 \times 10^5$ per ml. Count of ultra-violet irradiated suspension before illumination was $<10$ per ml.

† Exp. 2: Count of non-ultra-violet irradiated control was $2.2 \times 10^6$ per ml. Count of ultra-violet irradiated suspension before illumination was $<10$ per ml.

In another experiment (with different light source) a 3-fold recovery was observed after as little as 2 minutes of illumination, and 810-fold after 4 minutes. An experiment in which the duration of illumination was constant, but the intensity varied, showed that the rapidity of recovery was proportional to intensity, within limits.

*Temperature.*—In subsequent experiments there was employed a uniform artificial light source consisting of a slide projection lamp containing a 500-watt Mazda projection bulb. The outer lens of the lamp was placed about 5 cm. from the cells, in order to obtain as intense an illumination as possible. A conidial suspension was irradiated $1^1/_2$ minutes at 20 cm. from the ultra-violet lamp. Table 3 shows the effect of temperature on the rapidity of recovery, the visible light illumination being kept constant at 10 minutes. An independent ultra-violet irradiation was made for each temperature determination; this may partially account for some of the variability in the results. It is seen that the rate of recovery increases with rise in temperature up to about 50°C.

Ultra-violet irradiated suspensions could be kept at 5°C. in the dark for up to 4 hours without interfering with subsequent recovery when illuminated.

The knowledge furnished by the experiments just described enabled us to induce over 100,000-fold recovery with a high degree of reproducibility, by illuminating ultra-violet irradiated suspensions with a light source as described for 20 to 30 minutes at 37°C.

The light source used by us emitted infra-red as well as visible light. Since considerable work has been done on the effect on mutation and chromosomal rearrangements of pre- and posttreatment of x- or ultra-violet irradiated cells with near infra-red,[6, 7, 8] it was of importance to determine the comparative effect on recovery of the infra-red and visible components of our light source. Suspensions illuminated with light in which the infra-red had been eliminated by a filter[9] consisting of a 3.2-cm. deep cell containing 0.5 $N$ $CuCl_2$ aqueous solution, recovered almost as much as controls with no filter. This filter absorbs some of the visible red, as well as the infra-red. On the other hand, interposition of a filter consisting of a 3.2-cm. deep cell containing a saturated solution of $I_2$ in $CCl_4$, which eliminates most of the visible light and passes the infra-red[8, 9] resulted in no recovery at all. There was moderate recovery when an $I_2$-$CCl_4$ filter 1 cm. deep was used, but use of this filter was not a critical test, for a considerable portion of the visible light passed through this filter. These simple experiments do not of course exclude the possibility that infra-red illumination of sufficient intensity will not induce recovery; they do show that the most active component of our light source was the visible light. One of the main features of the infra-red-ultra-violet, or -x-ray studies,[6, 7, 8] is that pretreatment with infra-red has a marked effect on the behavior of cells to subsequent irradiation with ultra-violet or x-rays. We therefore illuminated conidial suspensions of *S. griseus* with visible light before irradiating with ultra-violet. There was no increase whatever in the survival rate on subsequent irradiation with ultra-violet.

The magnitude of the recovery phenomenon made it imperative to make sure that it was not due to some experimental artifact, such as declumping of clumped cells; and to ascertain whether the effect of visible light was on the menstruum rather than on the cells themselves.

Elimination of clumping and declumping as a factor was shown by experiments where ultra-violet and subsequent visible light irradiation was done on cells which had first been smeared on the surface of nutrient agar plates. Light-induced recovery occurred as usual.

That recovery was not due to a stimulation of germination in cells which had a long lag phase due to ultra-violet irradiation, was shown by the fact that prolonged incubation of plates which had been seeded with irradiated cells never disclosed the presence of slow-growing colonies. The maximum number of colonies was always reached after 3 days of incubation.

There was a possibility that the killing effect of ultra-violet light on *S. griseus* was due chiefly to ozone dissolved in the menstruum from the air, or to peroxides or other compounds formed in the menstruum by the ultra-violet light. If these toxic compounds rendered cells non-viable, then their elimination by decomposition by visible light, might allow cells to germinate and form colonies—i.e., recover.

Numerous experiments were made to detect a possible unusual sensitivity of *S. griseus* to the ultra-violet irradiated menstruum, with negative results. Air from the vicinity of the mercury lamp was bubbled for one hour through a suspension of cells, with no sign of toxicity. Sterile nutrient agar plates were irradiated for one hour, then inoculated with spores with no sign of more than a negligibly lower count than controls. Non-irradiated spores were added to suspensions of irradiated spores to see whether substances given off by irradiated cells might be toxic to non-irradiated cells with negative results. Any toxicity that was observed in these experiments never resulted in more than about 20 per cent killing, whereas ultra-violet irradiated cells under the conditions of our recovery experiments had usually a survival of the order of $1 \times 10^{-6}$.

*Discussion.*—The evidence presented suggests that in visible light we have a factor which uniformly, reproduceably causes the recovery of many of the cells which had been rendered non-viable by ultra-violet irradiation. The action is probably directly on the cells rather than on the menstruum, and there was no evidence of any experimental artifacts being involved. The magnitude of the effect makes it likely that a key factor in the lethal effect of ultra-violet light is being affected by the visible light. Whether or not light-induced recovery bears a relation to other types of recovery previously recorded is difficult to say. All such studies, as well as studies on ultra-violet induced mutation must be evaluated on the basis of whether light-induced recovery has played a part. There can be no doubt that the latter is at least partly responsible in some cases for the notorious variability of ultra-violet-mutation studies.

That the phenomena described here are not confined to actinomycetes only is suggested by observations in the older literature (summarized in the review by Prat[10]) of the antagonism to ultra-violet light of radiations of other wave lengths. These observations were usually made on cells or tissues irradiated by a mixture of wave lengths as compared to monochromatic irradiations, but consistently showed that the biological effect of ultra-violet light was diminished by simultaneous irradiation with visible or infra-red light. Since such effects were usually slight,[11] these older experiments are hard to evaluate. They, as well as other chemical and physical evidence of antagonism of ultra-violet and other light (also summarized by Prat[10]), suggest the phenomenon may be a general one.

While it is premature to do more than speculate on the mechanism involved in light-induced recovery, the following is suggested as a working hypothesis. Much of the killing effect of ultra-violet light is due to a light-labile alteration of some constituent in the cell. Exposure to visible light restores this altered constituent to its former state.

The powerful action of light on the resuscitation of the ultra-violet treated cell leads us to hope that further study of this phenomenon may yield clues leading to the discovery of factors causing similar recovery from x-irradiation of irradiation from radioactive materials. There is thus the possibility of at least a partial physiotherapy of radiation injury.

Of great importance is the relation of recovery to the mutagenic action of ultra-violet light. Work is in progress on light-induced recovery in the various microbial groups, such as bacteria, yeasts, fungi, and bacteriophage, and on the genetic aspects of light-induced recovery in microorganisms.

*Summary*.—Illumination with visible light will induce the recovery or the regaining of viability of ultra-violet irradiated conidia of the actinomycete, *S. griseus* ATC 3326. The light-induced recovery phenomenon is reproduceable and uniform and results in as high as a 400,000-fold increase in number of survivors in an ultra-violet irradiated suspension. The characteristics of the phenomenon are described, and its significance discussed.

---

* This study was aided by a grant from Schenley Laboratories, Inc.

[1] Hollaender, A., and Emmons, C. W., *Cold Spring Harbor Symp. on Quant. Biol.*, **9**, 179–186 (1941).

[2] Cook, E. V., *Radiology*, **32**, 289 (1939).

[3] Latarjet, R., *C. r. Acad. Sc. Paris*, **217**, 186–188 (1943).

[4] Kelner, A., *J. Bact.*, in press (1949).

[5] Kelner, A., *Ibid.*, **56**, 457–465 (1948).

[6] Kaufmann, B. P., and Gay, H., PROC. NAT. ACAD. SCI., **33**, 366–372 (1947).

[7] Swanson, C. P., Hollaender, A., and Kaufmann, B. N., *Genetics*, **33**, 429–437 (1948).

[8] Kaufmann, B. P., Hollaender, A., and Gay, H., *Ibid.*, **31**, 349–367 (1946).

[9] Brackett, F. S. *in* Duggar, B. M., *Biological Effects of Radiation*, McGraw-Hill, New York, 1936.

[10] Prat, S., *Protoplasma*, **26**, 113–149 (1936).

[11] Schreiber, H., *Strahlenther.*, **60**, 518–523 (1937).

# THE MODE OF ACTION OF THE METAL-PEPTIDASES*

By Emil L. Smith

Laboratory for the Study of Hereditary and Metabolic Disorders, and the Departments of Biochemistry and Medicine, University of Utah School of Medicine, Salt Lake City

Communicated by Henry Eyring, December 30, 1948

In 1926, Von Euler and Josephson[1] found that while a peptidase preparation from intestinal mucosa readily hydrolyzed glycylglycine, the N-acylated peptide was resistant to the enzyme. They concluded that a free amino group was required for the action of the enzyme and proposed "the diaffinity theory" which postulates that the hydrolysis of peptides takes place by a combination of the enzyme with the amino group of the substrate and at the peptide linkage to be split. With the later classification of dipeptidases, aminopeptidases and carboxypeptidases, the diaffinity theory (or polyaffinity theory) was extended to these enzymes, and, as the nomenclature suggests, the free polar groups required in the substrate were identified as essential for binding by the enzyme.

This simple scheme which was developed about 20 years ago has required two important modifications. First, it has now been conclusively demonstrated that the peptidases require for their action more than free polar groups in the substrate; that is, there are specific side-chain configurations which are also essential for the enzymatic action.[2] Thus, there has been identified leucine aminopeptidase, glycylglycine dipeptidase, etc. The second, and for present considerations the more important development, is that most of the well-characterized exopeptidases have been shown to be metalloproteins[3, 4, 5, 6] in which the metal is usually quite specific and essential for the enzymatic action.

In this paper, we propose to show how the metal functions in forming the enzyme-substrate complex, and why the formation of this complex must occur at two or more points for the hydrolytic action of the peptidases. While the theory of enzymatic action developed here has been considered primarily for certain peptidases, it undoubtedly will be applicable to other enzymes particularly those which exert a hydrolytic action.

*Theory of Peptidase Action.*—The theory to be considered rests on three primary postulates: first, that the formation of an enzyme-substrate complex ("the activated complex") consists in the formation of a compound with definite chemical bonds; second, that in the metal-containing peptidases, at least two of these bonds between the substrate and the enzyme are formed through the metal; additional linkages or points of attraction may exist through the metal or directly between the substrate and the protein moiety. Third, the two linkages formed between the metal and the substrate must be on opposite sides of the peptide bond. These two linkages (together with those which involve the protein directly), because of their great strength, produce an electronic distortion at the peptide bond; this labilization (decrease in energy of activation) permits the usual catalytic hydrolysis of the linkage by hydrogen and/or hydroxyl ions. It will be shown that there is considerable evidence which supports these postulates.

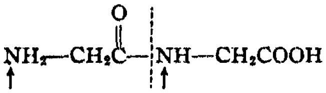
*The Enzyme-Substrate Complex.*—Ever since the promulgation of the Michaelis-Menten concept, considerable kinetic data have been amassed which demonstrates that the enzyme must form a complex with its substrate. In recent years, direct physical evidence has been obtained which shows that a definite chemical bonding takes place. The observation of Keilin and Mann[7] that peroxidase forms a complex with hydrogen peroxide which can be observed spectroscopically has been used by Chance[8] to elucidate the mechanism of this reaction. It has now also been demonstrated by Chance[9] that catalase likewise forms a definite intermediary complex with hydrogen peroxide. Thus, for two enzymes which are metalloproteins, the spectroscopic and other evidence indicates a binding of the substrate by the iron in the prosthetic group.

That the binding occurs through definite chemical bonds is also evident from the properties of most of the metal-complexes which are formed with simple substances. It is well known that these "Werner complexes" are extremely stable and show little or no tendency to ionize in aqueous solution. The soluble compounds formed by calcium or magnesium with polycarboxylic acids are also a case in point. A striking, though not unexpected, demonstration of this is given in a recent paper by Neuberg and Mandl[10] where it is reported that in the presence of amino acids, many metals such as Zn, Cu, Mn, Fe, etc., cannot be precipitated as insoluble sulfides. Thus, the complexes formed by the amino acids and the metals must consist of firm chemical bonds. The implications for complex formation of peptides and proteins with metals is obvious. Here, where chelate rings can be formed, a greater stability is possible. This is particularly evident with divalent cobalt where the most stable complexes are with compounds in which five-membered rings are formed.

*Linkage of Substrate and Enzyme through the Metal.*—In many peptidases and other enzymes, the metal is loosely and reversibly bound to protein since the metal may be removed by dialysis or certain purification procedures. The recombination of metal and protein has been shown to follow the usual mass law equilibrium.[4] With many enzymes the restoration of the full activity by the addition of metal ions is a slow reaction.[3] Obviously, complex formation of the metal and the protein must be occurring. This has been found with leucine aminopeptidase, prolidase, glycyl-L-leucine dipeptidase, and carnosinase; it has also been observed for the non-proteolytic enzymes, arginase and phosphatase.[11] Recently, it was found that pancreatic carboxypeptidase is a metal-protein as shown by its inhibition with the usual metal poisons. In this enzyme, the metal, which is presumably magnesium, is so firmly bound that it cannot be separated from the protein by any mild procedure.[12]

With all these peptidases, removal of the metal or combination of the metal with an inhibiting agent prevents the enzyme activity. Thus, the metal is absolutely essential for the enzyme action. The vague concept of "activation by metal ions" should be replaced by the full recognition of these enzymes as metalloproteins.

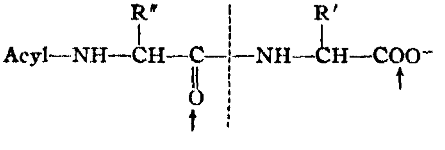
The requirement of the exopeptidases for substrates which contain free polar groups suggests that these polar groups combine, through complex formation, with the metal of the enzyme. It has already been indicated that with glycylglycine dipeptidase which requires cobalt for its activity, there is a marked parallelism between the ability of the enzyme to hydrolyze the substrate and the ability of $Co^{++}$ to form a coordination compound with the substrate.[5, 6] $Co^{++}$ combines with glycylglycine through the nitrogen of the free amino group and the nitrogen of the peptide bond as indicated by the resistance of dimethylglycylglycine and glycylsarcosine to hydrolysis by the enzyme, and by the failure of these compounds to coordinate with $Co^{++}$. A detailed description of these results has already been presented. It cannot be decided from the present evidence whether the essential free carboxyl group combines with the enzyme through the protein or through the metal. Nevertheless, it must be emphasized that the metal combines with at least two points on the substrate, and that these are on opposite sides of the sensitive bond. The



$$\underset{\uparrow}{NH_2}\text{—}CH_2\overset{\overset{O}{\|}}{C}\text{—}\vdots\underset{\uparrow}{NH}\text{—}CH_2COOH$$

arrows indicate the points of combination; the dotted line is where cleavage occurs. Compounds like glycine which are unable to form a chelate complex with $Co^{++}$ form a much weaker compound as indicated by the much smaller influence of glycine on the absorption spectrum of $Co^{++}$ as compared with the complex of $Co^{++}$ and glycylglycine. The weak combination of glycine with $Co^{++}$ suggests that once hydrolysis of the peptide bond occurs, the two glycine fragments will readily be displaced by the dipeptide for which the $Co^{++}$ has a higher affinity.

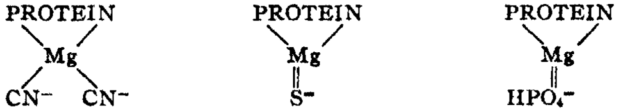
With carboxypeptidase, a somewhat different situation occurs. Here magnesium is the metal. The enzyme is inhibited not only by cyanide, sulfide and cysteine, but also by agents which are fairly specific in binding with magnesium such as citrate, oxalate, pyrophosphate and orthophosphate.[15] The substrates of carboxypeptidase do not contain a free amino group but do have a free carboxyl group. Moreover, a peptide hydrogen is not required as indicated by the slow enzymatic splitting of carbobenzoxy-L-tryptophyl-L-proline and, even more dramatically, by the rapid hydrolysis of an ester bond in the compound hippurylphenyllactic acid.[15] With this enzyme, chelate complexing of the magnesium must occur through the carbonyl group (or its enol form) at the sensitive bond and with the ionized carboxyl group, as shown for an acylated dipeptide.



$$\text{Acyl—NH—}\overset{\text{R}''}{\overset{|}{\text{CH}}}\text{—}\underset{\uparrow}{\underset{\text{O}}{\underset{\|}{\text{C}}}}\text{—}\vdots\text{—NH—}\overset{\text{R}'}{\overset{|}{\text{CH}}}\text{—}\underset{\uparrow}{\text{COO}^-}$$

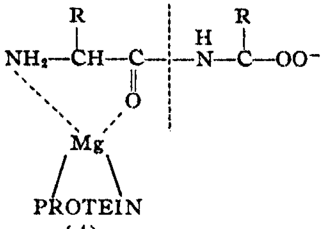
Again, the arrows indicate the points of metal combination and the dotted line the sensitive bond. In contrast to aminopeptidases, glycylglycine dipeptidase and probably other dipeptidases, the metal in carboxypeptidase does not combine with nitrogen at all.

That a chelate complex is formed by the magnesium of carboxypeptidase and the substrate is suggested by two lines of evidence. First of all, magnesium is known to form strong complexes of chelate character with dicarboxylic acids and other compounds. Secondly, quantitative studies of the inhibition of carboxypeptidase show that for each active enzyme center, there is combination with one orthophosphate, one sulfide, or two cyanide ions.[18] Thus, with this metal which can form four bonds, two coördinate and two of covalent character, there are two valences available for combination with the inhibitor, and presumably two are strongly linked with the protein. Carboxypeptidase apparently contains sulfhydryl groups as shown by the strong inhibition of the enzyme produced by iodoacetate and by traces of heavy metals such as $Cu^+$ and $Pb^{++}$. It is reasonable to suppose that in the absence of substrate or inhibitor, two valences of the magnesium are tightly bound to unknown groups of the protein, and two are loosely bound to sulfhydryl groups. In the presence of inhibiting ions, the two loosely bound valences are reversibly bound to the inhibitor. In the presence of a suitable substrate, the loose combinations with the protein are displaced by the substrate for which the metal must have a much greater affinity. It is apparent that this hypothesis for carboxypeptidase is somewhat similar to the explanation of the properties of hemoglobin where of the six available bonds of the iron, four are coördinated to the nitrogens of the pyrrol rings, one is tightly bound to imidazole nitrogen, and one is loosely bound to imidazole; it is the last



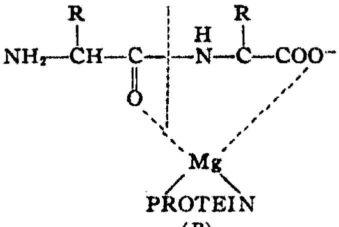
linkage which is displaced by molecular oxygen or carbon monoxide. It is unnecessary to recapitulate here the excellent evidence which has been adduced in developing our knowledge of the properties of hemoglobin.[16] Nevertheless, this oxygen-binding protein serves as a model for the completely analogous situation which must occur with the metal-enzymes.

Carboxypeptidase hydrolyzes carbobenzoxyglycyl-L-tyrosine about 20,000 times more rapidly than it does the dipeptide glycyl-L-tyrosine.[17] On the basis of our postulates, it is possible to explain why the free amino group renders the dipeptide so much more resistant to the hydrolytic action of the enzyme. Magnesium combines very readily with free amino groups, and functions as the metal in such enzymes as leucine aminopeptidase. With the dipeptide, carboxypeptidase will have a greater tendency to coördinate with the amino group and the adjacent carbonyl group as in *A* than with the carbonyl and the carboxyl as in *B*. It is evident that coördination with both the amino and carboxyl group as in *C* would give a compound in which no large electronic effect would be exerted at the peptide bond. It would be expected that an equilibrium,



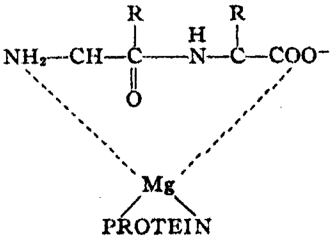
$\mathrm{NH_2{-}CH(R){-}C({=}O){-}N(H){-}CH(R){-}OO^-}$

(*A*)



$\mathrm{NH_2{-}CH(R){-}C({=}O){-}N(H){-}CH(R){-}COO^-}$

(*B*)



$\mathrm{NH_2{-}CH(R){-}C({=}O){-}N(H){-}CH(R){-}COO^-}$

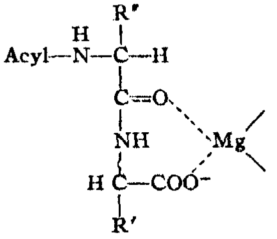
(*C*)

preponderantly in favor of *A*, would exist between the two complexes *A* and *B* (or *A*, *B* and *C*). According to our theory, there would be no hydrolytic action involved in *A* since the two bonds are on the same side of the molecule with respect to the sensitive peptide linkage. Only *B* would be subject to hydrolytic action since the structure is the same as that postulated for the combination of the acylated dipeptide with the enzyme.

Bergmann[18] and his collaborators have already shown how the polyaffinity theory helps to explain the optical specificity of the peptidases.

He postulated that "the polar groups of the enzyme act on the polar groups of the substrate and force the latter to assume a definite spatial position." With a compound containing an unnatural amino acid of the D configuration, it was assumed that the large side-chain "lies between the enzyme and the decisive atomic groups of the dipeptide and prevents the enzyme from approaching the substrate."

With the added knowledge which is now available that the binding of the substrate occurs at least in part through the polar groups of the substrate to the metal of the enzyme, it is possible to give a more precise formulation of the optical specificity requirements of the substrates and of inhibitors of carboxypeptidase. This enzyme has strong points of combination not only through the metal to the polar groups of the substrate but also with the side-chain of the amino acid residue which bears the free carboxyl group. The affinity of this side-chain is greatest for compounds which contain an aromatic ring, phenylalanine, tyrosine and tryptophane, less for those with large aliphatic residues, leucine and isoleucine, and still less for alanine and glycine.[19] We must then suppose a definite spatial relationship for the enzyme in which the substrate, an acylated dipeptide, containing L amino acids, must fit.

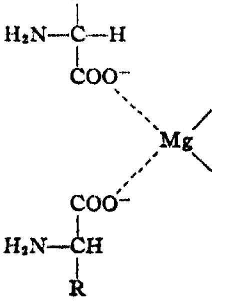


The outline indicates a schematic representation of the surfaces of the protein. There are at least four important points of attraction, the two valences of the metal which bind the essential carbonyl and carboxyl, and the two specific protein centers which attract R′ and R″. At the present time, the nature of these last two portions of the protein molecule is unknown. However, R″ is of minor importance, since variations in the nature of R″ have only a relatively small influence on the rate of hydrolysis of the compound. For example, the rate of hydrolysis of carbobenzoxy-L-tryptophylglycine is only three times the rate of hydrolysis of carbobenzoxyglycylglycine.[14] The tremendous influence of R′ is shown by comparison of the rate for carbobenzoxyglycyl-L-tryptophan which is about 2000 times greater than that of carbobenzoxyglycylglycine. The protein center which attracts R′ must, therefore, have a tremendous affinity for the side-chain R′. Compounds which contain D amino acids at either R′ or R″ are not hydrolyzed by the enzyme.[18, 19] Thus, the four known centers of attraction or binding in the enzyme must have a fixed spatial configuration.

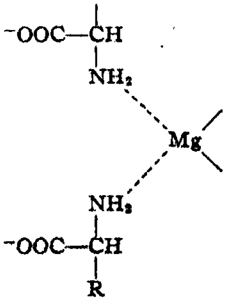
*The chelate ring which is formed by the magnesium helps to explain the high affinity of the enzyme for its substrates.* Compounds which do not form a chelate ring with the magnesium must have a much lower affinity with the enzyme. This explains why L amino acids do not possess any marked ability to inhibit this enzyme. It also explains the high turnover rate of this and other peptidases. Since the products of the hydrolytic reaction do not form chelate complexes, they have a poor affinity for the enzyme and are rapidly released after the peptide bond is split.

It is of great interest that while L amino acids and an acylated dipeptide such as carbobenzoxy glycyl-D-phenylalanine are not inhibitors of the enzyme, D amino acids do inhibit.[17] An explanation of this is apparent from the structure shown above. The acylated dipeptide which contains a D amino acid cannot fit into the structure of the protein in such a manner as to combine with the metal and with the two side-chain positions simultaneously. Such a compound will be able to fit the protein structure only at the side-chain positions or with the metal. Therefore, the affinity of the enzyme for such a compound is very much smaller than the affinity for the isomeric compound containing L amino acids.

To explain the poor affinity of an L amino acid for the enzyme, we must suppose that the high affinity of the side-chain positions results in an orientation such that the carboxyl group is presented to the metal, as in the usual substrate. Since no chelation is possible, the affinity is low. However, with the D amino acid we can assume that when the side-chain position is fixed, the molecule is orientated so that the amino group can combine with the metal. This will resemble the high affinity combination found with the dipeptide, and we can expect that dipeptides of appropriate configuration will be strong competitive inhibitors. The following diagrams illustrate this explanation:



Orientation with L Amino Acid

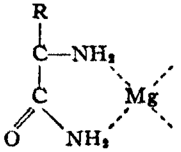


Orientation with D Amino Acid

*Hydrolysis of the Peptide Bond.*—It is apparent from what has already been presented that the present hypothesis is a form of the "strain or deformation theory" first introduced into physical chemistry many years ago. It is now possible to make an experimental approach to the mechanism of certain enzymatic reactions because enough is known concerning the binding properties of metals for the possible substrates of these enzymes. If the present theory is correct, then the decreased energy of activation of the peptide bond is produced by the attraction of electrons by the metal, through its combination with the protein, while the protein centers which attract or bind R′ and R″ serve to orient the substrate and fix it into position so that the strong electronic attraction of the metal produces an electronic deformation. The simplest explanation of the resulting hydrolysis of the peptide bond is that it is due to the usual hydrogen and/or hydroxyl ion catalytic effects. These ions, at high concentrations, do catalyze the scission of peptide bonds. At pH values near neutrality their effect is negligible, but with a greatly modified bond at which the electronic positions are altered requiring a low energy of activation, the reaction velocity becomes appreciable. The pH activity function of the enzyme gives no hint of this because it is a summation of these and other factors: ionization of the substrate, complexing of metal and substrate, ionic forces within the protein which help to determine the side-chain affinities, and the stability of the protein itself. A portion of this may be evaluated by the change in the pH activity curve which results when different metals are substituted with the same enzyme. This was first described by Hellerman and Stock for arginase.[20] The same effect has now been found for a peptidase, carnosinase,[11] which is active with either $Mn^{++}$ or $Zn^{++}$ but where the resulting pH curves are different.

There is considerable evidence which indicates that an ionic catalysis is involved in the peptide bond hydrolysis. This is shown by the tremendous influence on the sensitivity of the peptide bond of the strength of the acid which is combined at the sensitive linkage. This effect is now being studied quantitatively and in considerable detail. Qualitatively, the effect is already known although the facts have not previously been interpreted in this manner. For example, chloroacetylamino acids are much more sensitive to hydrolysis than acetylamino acids; chloroacetic acid is a much stronger acid than acetic acid. Similarly, it was recently found that carbobenzoxy β-alanylamino acids are hydrolyzed much more slowly than carbobenzoxyglycylamino acids.[21] Again it is well known that the alpha amino carboxylic acids are much stronger acids than beta amino carboxylic acids. It has also been found[22] that carbobenzoxyamino acids are hydrolyzed more slowly than carbobenzoxydipeptides where the same relationship may be expected. These facts are completely in agreement with the classical knowledge that the sensitivity to hydrolysis of esters and amides is proportional to the strength of the acids which are linked in these compounds.

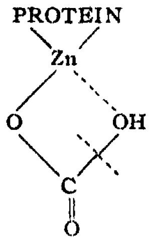
The preceding discussion has dealt so far with a dipeptidase and a carboxypeptidase. With an aminopeptidase, the metal binding must be similar to that in the dipeptidase. With leucine aminopeptidase, either $Mg^{++}$ or $Mn^{++}$ can function as the metal. The complex will then form with the free amino group and with the amide nitrogen. The side chain, R, is bound directly by the protein. The diagram illustrates the type of

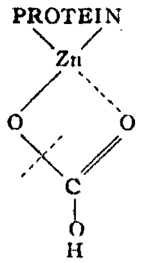


complex to be expected, and the resulting action would be the same as that already discussed.

The present hypothesis of peptidase action appears to be directly applicable to a vast number of enzymatic reactions. Kornberg, Ochoa, and Mehler[23] have shown that in some instances β-keto acids form unstable complexes with metal ions which result in decarboxylation. Here the electronic deformation is sufficient to permit the ionic catalysis to proceed. In other cases, the metal forms a stable complex with the keto acid; addition of the specific protein then leads to decarboxylation. All of these complexes are observed spectroscopically, and as with the peptidases, the complexes must be of chelate character. These investigators have suggested that the role of the enzyme in oxalosuccinic decarboxylation is to catalyze the formation of the substrate-$Mn^{++}$ complex. According to our hypothesis, the metal-substrate complex is bound to protein in such a manner as to produce an electronic deformation at the sensitive bond. The real catalysis is then an acid or base catalysis by the mechanism proposed by Bronsted. It is striking that acylated amino acids or peptides are also keto acids and that while the metal combination is similar to that in the β-keto acids, the resulting enzymatic action on the compound is quite different.

Perhaps the simplest example of our theory concerns its application to carbonic anhydrase, a zinc enzyme.[24] The probable substrate at neutral pH values is bicarbonate ion. Two of the four possible bonds of zinc are with the protein; the other two must bind with the bicarbonate to form a complex as shown below:





Acid or base catalysis at the sensitive point, indicated by the broken line, causes the decomposition of the compound which results in weaker linkages with the zinc; now one bond will be loosely combined with CO$_2$ and the other with hydroxyl ion. Since the action of carbonic anhydrase is reversible, this must be the binding that leads to the formation of bicarbonate ion. It should be recalled that zinc forms an extremely insoluble (poorly ionized) carbonate, as well as a subcarbonate which contains hydroxyl groups. Thus, these linkages are identical with those postulated for the enzyme-substrate complex.

Extension of this hypothesis to other types of metal-enzyme reactions is obvious and will not be pursued at this time. The present theory may, of course, be partially or completely incorrect. However, it is felt that it does have the merit of being subject to further experimental test. Moreover, it provides the first comprehensive explanation for many previously unexplained and apparently unrelated observations. It certainly helps to move much that has been mysterious in enzymatic reactions into the sphere of ordinary catalytic mechanisms, and may serve to elucidate the mechanism of certain non-enzymatic reactions also. However, the role of the protein in determining most of the enzymatic specificity remains a problem for the future.

* Supported by a grant from the United States Public Health Service.

[1] Von Euler, H., and Josephson, K., *Z. physiol. Chem.*, **157**, 122 (1926).

[2] Bergmann, M., and Fruton, J. S., in Nord, F. F., and Werkman, C. H., *Advances in Enzymology and Related Subjects, New York*, **1**, 63 (1941); Bergmann, M., *Ibid.*, **2**, 49 (1942).

[3] Smith, E. L., and Bergmann, M., *J. Biol. Chem.*, **138**, 789 (1941); **153**, 627 (1944).

[4] Smith, E. L., *J. Biol. Chem.*, **163**, 15 (1946).

[5] Smith, E. L., *Ibid.*, **173**, 553, 571 (1948).

[6] Smith, E. L., *Ibid.*, **176**, 21, 39 (1948).

[7] Keilin, D., and Mann, T., *Proc. Roy. Soc. London*, **B122**, 119 (1937).

[8] Chance, B., *J. Biol. Chem.*, **151**, 553 (1943).

[9] Chance, B., *Acta Chem. Scand.*, **1**, 236 (1947).

[10] Neuberg, C., and Mandl, I., *Arch. Biochem.*, **19**, 149 (1948).

[11] This work has been reviewed[6] except for carnosinase (Hanson, H. T., and Smith, E. L., to be published).

[12] Smith, E. L., and Hanson, H. T., *J. Biol. Chem.*, **176**, 997 (1948).

[13] Smith, E. L., and Hanson, H. T., to be published.

[14] Smith, E. L., *J. Biol. Chem.*, **175**, 39 (1948).

[15] Snoke, J. E., Schwert, G. W., and Neurath, H., *J. Biol. Chem.*, **175**, 7 (1948).

[16] Wyman, J., Jr., *Federation Proc.*, **7**, 502 (1948).

[17] Calculated from data of Hofmann, K., and Bergmann, M., *J. Biol. Chem.*, **134**, 225 (1940).

[18] Bergmann, M., *Harvey Lectures*, **31**, 37 (1935–1936).

[19] Stahmann, M. A., Fruton, J. S., and Bergmann, M., *J. Biol. Chem.*, **164**, 753 (1946).

[20] Hellerman, L., and Stock, C. C., *Ibid.*, **125**, 771 (1938).

[21] Hanson, H. T., and Smith, E. L., *Ibid.*, **175**, 833 (1948).

[22] Bergmann, M., and Fruton, J. S., *Ibid.*, **145**, 247 (1942).

[23] Kornberg, A., Ochoa, S., and Mehler, A. H., *Ibid.*, **174**, 159 (1948).

[24] Keilin, D., and Mann, T., *Biochem. J.*, **34**, 1163 (1940).

---

# *FURTHER EXPERIMENTS WITH VITAMIN A IN RELATION TO AGING AND TO LENGTH OF LIFE**

By Henry C. Sherman and Helen Yarmolinsky Trupp

Department of Chemistry, Columbia University

Communicated December 17, 1948

Previous work in this department[1] has shown that, when a minimal-adequate allowance of vitamin A was doubled, the length of life of the experimental animals (rats) was significantly increased: males, about 5 per cent; and females, about 10 per cent. These females showed, in the same experiments, an increase of about 17 per cent in the duration of reproductive life—an objective criterion sometimes taken as symbolizing "prime of life" or "length of useful life."

TABLE 1

Influence of Vitamin A Value of Food upon Aging and Length of Life of Rats: Number of Cases in Each Average ( )

| | ON DIET 361, 12 I. U./G., 1942–1948 | ON DIET 361, 12 I. U./G., 1945–1948 | ON DIET 362, 24 I. U./G., 1945–1948 |
|---|---|---|---|
| Reproductive life of females | (72) 378 days | (36) 387 days | (36) 292 days |
| Length of life: | | | |
| Of females | (72) 812 days | (36) 794 days | (36) 777 days |
| Of males | (48) 720 days | (24) 716 days | (24) 647 days |

In those experiments the amount of vitamin A fed as minimal-adequate was 3 I. U. per gram of air-dry food mixture, or about 0.8 I. U. per calorie. Both at the 3 I. U. and at the 6 I. U. level the numbers of cases (over 100 males and over 150 females) were large enough to be conclusive. A smaller number of experiments suggested that an intake level of 12 I. U. per gram might be still more beneficial.

The present paper reports results of further experiments at 12 I. U., and also parallel experiments at the still higher level of 24 I. U. per gram of air-dry food.

The numerical data are summarized in table 1, in which the first column of figures shows the combined results of both the 1942–1945 and the 1945–1948 series at the level of 12 I. U. per gram (Diet 361). As the data of these two series are in good agreement, their combined averages constitute the most conclusive evidence for this level of feeding, and link the present series with those published in these PROCEEDINGS in 1945.[1]

The second and third columns of figures in the present table 1 summarize the data of the side-by-side comparisons of Diet 361 with 12 I. U., and Diet 362 with twice as much vitamin A per gram. Thus, without repetition of data, the findings of the two series are so linked as to permit numerical comparisons of the final respective averages at the four levels studied, namely, 3, 6, 12 and 24 I. U. per gram of air-dry food. Such combination and comparison of the 1942–1945 and the 1945–1948 data of this laboratory show that:

(1) The intake level of 3 I. U. of vitamin A per gram of dry food, or about 0.8 I. U. per calorie, meets a standard of minimal adequacy. Human subjects have subsisted at such levels for many months without showing signs of deficiency, and a laboratory rat family is thriving in the 67th generation at this level.

(2) Yet while such a level of vitamin A intake thus clearly meets a certain minimal standard of adequacy it does not, in the long-term test of a lifetime, support the optimal life history of which the individual is capable, for with double the allowance (rats on Diet 360, with 6 I. U. of vitamin A per gram of dry food) the length of life was increased about 5 per cent in males and about 10 per cent in females.

(3) On again doubling the level of intake of vitamin A there was again a measurably increased benefit making the averages compared with those of the 3 I. U. level, an increase in length of life of about 10 per cent in males and about 12 per cent in females.

(4) At both the 6 I. U. and the 12 I. U. levels, reproductive life in the females—the best strictly objective measurement we know by which to symbolize the postponement of aging or increase of "useful" life—was increased in larger proportion than the length of life itself. Obviously the newer knowledge of nutrition is still too young to have made similarly controlled studies of entire lives of human subjects, but what we know of the chemistry of the nutrition of the human and rat species makes it probable that the same principle applies and that increasing liberality of intake of vitamin A brings increasing benefit to nutritional well-being and length of life up to higher levels than previously supposed.

(5) As judged by sufficiently long-term and closely controlled experi-

ments, optimal intakes of vitamin A appear to be at least three to four times as high as the levels of minimal adequacy sometimes accepted. Doubtless, too, optimal level of intake is less a point than a plateau. Thus in the experiments here discussed, the plateau of optimal nutritional response extended from somewhere between 6 and 12 I. U. to somewhere between 12 and 24 I. U. per gram of food.

(6) At some points, but not all, the responses were less favorable in the animals fed at the 24 I. U., than in those fed at the 12 I. U., per gram, level of intake of vitamin A.

Fuller comparisons of the nutritional responses to the 12 and the 24 I. U. per gram levels may best be taken up in connection with a more detailed presentation of the data elsewhere.

* Aided by grants from The Nutrition Foundation, Inc., to Columbia University.

[1] Sherman, H. C., Campbell, H. L., Udiljak, M., and Yarmolinsky, H., PROC. NAT. ACAD. SCI., **31**, 107 (1945).

# ON DIFFERENCE SETS

BY S. CHOWLA

INSTITUTE FOR ADVANCED STUDY, PRINCETON, NEW JERSEY

Communicated by Oswald Veblen, December 5, 1948

It is known that the existence of $m + 1$ integers (called a perfect difference set of order $m + 1$) $d_1, d_2, \ldots, d_{m+1}$ such that the congruence $d_i - d_j \equiv n (\text{mod } m^2 + m + 1)$ has exactly one solution for every $n \not\equiv 0 (\text{mod } m^2 + m + 1)$, leads to the construction of a finite projective plane with $m + 1$ points on a line. All known cases of such planes have $m = p^g$, where $p$ is a prime and $g$ is a positive integer. It seems natural to conjecture that a perfect difference set of order $m + 1$ can exist only if $m$ is a power of a prime. Only recently Bruck and Ryser (*Bull. Am. Math. Soc.*) announced the startling result that there exists no finite projective plane with $m + 1$ points on a line whenever $m \equiv 1$ or $2 (\text{mod } 4)$, provided $m$ is divisible by a prime $4k + 3$ to an odd power. It follows from their result that for such values of $m$ there exists no perfect difference set (P.D.S.) or order $m + 1$. Moreover Singer proved the existence of a P.D.S. of order $m + 1$, whenever $m = p^g$.

In this paper I obtain some (but not all) assertions on the non-existence of perfect difference sets implied by the results of Bruck and Ryser. I also obtain some results not implied by their work. For example, I prove that there exists no P.D.S. of order $m + 1$ when $m = 10$ or $159$.

We introduce the idea of a difference set (D.S.) of $m$ numbers (mod $g$). We call the set of $m$ numbers $d_1, d_2, \ldots, d_m$ a difference set (mod $g$) if the congruence $d_i - d_j \equiv n (\text{mod } g)$ has the same number of solutions [which must be $m(m-1)/(g-1)$] for every $n \not\equiv 0 (\text{mod } g)$. The set of 5 numbers 1, 3, 4, 5, 9 furnishes an example of a D.S. of 5 numbers (mod 11). Again the set of 4 numbers 0, 1, 3, 9 forms a D.S. (mod 13). [This set occurs in Veblen and Bussey, *Trans. Am. Math. Soc.*, 1906, and is used to generate a finite projective plane with 4 points on a line.]

We now prove

THEOREM 1. *Let $m$ and $g$ be positive integers such that $m(m-1) \equiv 0 (\text{mod } \overline{g-1})$. Write $\theta = m - m(m-1)/(g-1)$. Let $g$ contain a prime factor $\lambda \equiv 3 (\text{mod } 4)$ such that $-\lambda$ is a quadratic non-residue of some prime factor $\phi$ of $\theta$, where $\phi$ occurs in $\theta$ to an odd power. Then there exists no D.S. of $m$ numbers (mod $g$).*

*Proof:* If possible let there exist a D.S. of $m$ numbers (mod $g$), say the numbers $d_1, d_2, \ldots, d_m$. Consider the sum

$$S = \sum_{j=1}^{m} \rho^{d_j},$$

where $\rho = \exp(2\pi i/\lambda)$. Clearly $S$ is an algebraic integer of the field $K(\rho)$. Further, since the $d$'s form a difference set (mod $g$), we have

$$S\bar{S} = m + \frac{m(m-1)}{(g-1)} \{\rho + \rho^2 + \rho^3 + \ldots + \rho^g\}$$

$$= m - \frac{m(m-1)}{(g-1)} = \theta,$$

$$S\bar{S} = \theta \qquad (1)$$

Now it is implicit in the theory of cyclotomy (as developed by Gauss) that the norm of $S$ in the field generated by $\rho$ is an integer of the form $(u^2 + \lambda v^2)/4$, where $u$ and $v$ are integers (we here use the fact that $\lambda$ is of the form $4k + 3$). On the other hand it follows from (1) that this norm is also equal to $\theta^{(\lambda - 1)/2}$. Hence we have

$$u^2 + \lambda v^2 = 4 \cdot \theta^{(\lambda - 1)/2}.$$

Since $(\lambda - 1)/2$ is odd, it follows from the last equation that $-\lambda$ is a quadratic residue of $\phi$ contrary to our assumption. Our theorem is proved.

*Examples:* 1. There exists no P.D.S. of order 7; i.e., there exists no D.S. of 7 numbers (mod 43). [Problem proposed by Veblen in *Am. Math. Monthly*, **13,** 46 (1906).]

*Proof:* Here $g = \lambda = 43$, $\theta = 7 - \frac{4^2}{4^2} = 6$; take $\theta = 3$. Clearly $-\lambda$ is a quadratic non-residue of $\phi$.

2. There exists no P.D.S. of order 11, i.e., there exists no D.S. of 11 numbers (mod 111). This result is new.

*Proof:* Here $g = 111$, $\lambda = 3$, $\theta = 10$. Take $\phi = 5$. Clearly $-\lambda$ is a quadratic non-residue of $\phi$.

3. There exists no P.D.S. of order 22 (Bruck and Ryser), i.e., no D.S. of 22 numbers (mod 463).

*Proof:* Here $g = \lambda = 463$, $\theta = 22$. Take $\phi = 11$. Clearly $-\lambda$ is a quadratic non-residue of $\phi$.

4. There exists no P.D.S. of order 160. This result is new.

*Proof:* We show that there cannot exist a D.S. of 160 numbers (mod $159^2 + 159 + 1$). Since $159^2 + 159 + 1 \equiv 0 \pmod{19}$, we may take $\lambda = 19$. Take $\phi = 3$ since $\theta = 159$. Clearly $-\lambda$ is a quadratic non-residue of $\phi$.

---

# CHARACTERS OVER CERTAIN TYPES OF RINGS WITH APPLICATIONS TO THE THEORY OF EQUATIONS IN A FINITE FIELD

By L. K. Hua and H. S. Vandiver

Department of Mathematics, University of Illinois, and Department of Applied Mathematics, University of Texas

Communicated December 17, 1948

Characters, modulo $m$, have been defined (for example by Landau[1]) and have been employed extensively in Group Theory and Number Theory, particularly in some of the most important parts of Analytic Number Theory. However, we may define them by means of any ring with a unity element and which contains only a finite number of elements which are not zero divisors, instead of the special ring consisting of the residue classes modulo $m$, and a number of extensions of known results concerning said special case are immediate. We will set up this concept and then specialize it for the finite field using results which are then applied to the theory of certain equations in a finite field which were considered in another paper[2] by the authors. This type of equation is

$$c_1x_1^{a_1} + c_2x_2^{a_2} + \ldots + c_sx_s^{a_s} + c_{s+1} = 0 \qquad (1)$$

where the $a$'s are integers such that $0 < a < p^n - 1$; $s \geqq 2$ for $c_{s+1} \neq 0$ and $s > 2$ for $c_{s+1} = 0$, the $c$'s being given elements of a finite field of order $p^n$, $p$ prime, which will be designated by $F(p^n)$ and

$$c_1 \ldots c_sx_1 \ldots x_s \neq 0. \qquad (2)$$

In the present paper we obtain a new explicit expression for the number of solutions of (1) in terms of generalized Lagrange resolvents and (see the $\tau$ number defined in Theorem I) and using this we find better limits for the number of solutions of (1) than those derived in E.

Let $R$ be a ring[2] with a unity element which has the property of containing only a finite number of elements which are not divisors of zero. A set of such elements forms a group. For, since it does not contain zero or any divisors of zero, the cancellation holds and a finite semi-group having this property is known to be a group. Or, otherwise expressed, these elements are units in $R$.

Let an arithmetical function $\chi(a)$ be defined for any element $a$ of $R$ so that

I. $\chi(a)$ is a complex number.
II. $\chi(1) \neq 0$.
III. $\chi(a) = 0$ when $a$ is zero or any divisor of zero.
IV. If $a_1 = a_2$ in $R$, then $\chi(a_1) = \chi(a_2)$.
V. $\chi(a_1a_2) = \chi(a_1)\chi(a_2)$.

The arguments employed by Landau[1] for the special ring formed by the residue classes modulo $m$ by which he obtains his Satz 127 to Satz 133 inclusive may be easily extended to yield the following:

For each character $\chi$ we have

$$\chi(1) = 1. \tag{3}$$

There is a character designated by $\chi_1$ such that

$$\chi_0(a) = 1, \tag{3a}$$

for each $a$ in $R$. Let $\phi(R) = h$ be the number of distinct units in $R$. We shall call $h$ the *indicator*[3] of $R$. Then since the units $u$ form a group, $u^h = 1$ in $R$, and by I and V,

$$\chi(a) = e^{2i\pi b/h}; \tag{4}$$

for $0 \leqq b < h$.

We also obtain

LEMMA I.

$$\sum_{a \in R} \chi(a) = \begin{cases} h \text{ for } \chi = \chi_0, \\ 0 \text{ for } \chi \neq \chi_0. \end{cases} \tag{5}$$

Suppose that we have, if $h \equiv 0 \pmod{k}$, a character such that $(\chi(a))^k = 1$ for each unit $a$ in $R$, this will be called a special character and designated by $\chi_k$. In particular $\chi_0$ is a special character for any $k$.

How let $R$ be a finite field of order $p^n$, $p$ prime, and designate the same by $F(p^n)$. Let $\theta$ be a primitive root of this field, and let $\beta$ be a primitive $(p^n - 1)$th root of unity. Then by (4)

$$\theta = \beta^b;\ 0 \leqq b < p^n - 1,$$

and since all the elements, $\neq 0$, of $F(p^n)$ are given by powers of $\theta$ then the whole character, by V, is given by powers of $\beta^b$. Now consider $\chi_k$. Then

$$\chi_k(\theta) = \beta^d \tag{6}$$

and by definition of $\chi_k$, $(\chi_k(\theta))^k = \beta^{dk} = 1$ which gives $d \equiv 0 \pmod{(p^n-1)/k}$. It follows from this that there are exactly $k$ special characters, and in particular there are $h$ distinct characters defined by $R$. We now have

LEMMA II.

$$\sum_{\chi_k} \chi_k(a) = \begin{cases} k \text{ for } a \text{ a } k\text{th. power,} \\ 0 \text{ otherwise.} \end{cases} \tag{7}$$

This follows easily from (6) and the remarks following it.

We also have[5]

LEMMA III. *If* $\zeta = e^{2i\pi/p}$, *then in* $F(p^n)$,

$$\sum_{a \epsilon k} \zeta^{tr(at)} = \begin{cases} 0 \text{ for } t \neq 0, \\ p^n \text{ for } t = 0. \end{cases} \tag{8}$$

We now proceed to[6]

THEOREM I. *The number of solutions* $N$ *of (1) with the restriction (2) is*

$$N = \frac{(p^n - 1)^s}{p^n} + \frac{T}{p^n}$$

*where*

$$T = \sum_{\substack{a \epsilon k \\ a \neq 0}} \prod_{i=1} \sum_{\chi ki} \chi_{ki}\,(a^{-1}c_i^{-1})\tau(\chi_{ki})\zeta^{tr(ac_s + 1)};$$

$$\tau(\chi_{ki}) = \sum_{b \epsilon k} \chi_{ki}(b)\zeta^{tr(b)}$$

*and* $k_i = (a_i, p^n - 1)$. *Also* $K$ *stands for* $F(p^n)$.

To prove this we note that

$$N = \frac{1}{p^n} \sum_{a \epsilon k} \sum_{i=1}^{s} \sum_{\substack{x_i \epsilon K \\ x_i \neq 0}} \zeta^{tr(aG(x))}; \tag{9}$$

$$G(x) = c_1x_2^{k_1} + c_2x_3^{k_2} + \ldots + c_sx_s^{k_s} + c_{s+1}.$$

First it is known that the number of solutions of (1) and the number of solutions of $G(x) = 0$ are the same if $(p^n - 1, a_i) = k_i$. Also when we carry out the summation in (9a) for the $x_i$'s all the terms are unity when we have a set of $x$'s which satisfy $G(x) = 0$ with the restriction (2) and when we sum with respect to $a$ we obtain $p^n$ according to Lemma III, but when we have a set of $x_i$'s such that (1) is not satisfied then by the same lemma the sum with respect to $a$ gives zero. (The relation (15) in E, p. 261, is a transformation of (9).) In the right-hand member, if we separate the zero and non-zero values of $a$, we obtain from (9),

$$N = \frac{(p^n - 1)^s}{p^n} + \frac{T}{p^n}, \tag{10}$$

where

$$T = \frac{1}{p^n} \sum_{\substack{a \in K \\ a \neq 0}} \sum_{i=1}^{s} \sum_{\substack{x_i \in K \\ x_i \neq 0}} \zeta^{tr(aG(x))}$$

Now employing Lemma II,

$$\begin{aligned} \sum_{\substack{x \in K \\ x \neq 0}} \zeta^{tr(cx^k)} &= \sum_{y \in K} \sum_{\chi_k} \chi_k\,(c^{-1}y)\zeta^{tr(y)} \\ &= \sum_{\chi_k} \chi_k(c^{-1})\tau(\chi_k) \end{aligned}$$

and then (10) gives

$$T = \sum_{\substack{a \in K \\ a \neq 0}} \prod_{i=1}^{s} \sum_{\chi_{k_i}} \chi_{k_i}\,(c_i{}^{-1}\,a^{-1})\tau(\chi_{k_i})\zeta^{tr(ac_{s+1})}. \tag{11}$$

We now employ (10) and (11) to find limits for $N$. We see from (4) that for a fixed $a$

$$\prod_{i=1}^{s} \chi_{u_i}(a) = \chi_v(a), \tag{11a}$$

for some $v$. Also it is known[6] (p. 152) that

$$\begin{aligned} |\tau(\chi)| &= p^{n/2}, \qquad \chi \neq \chi_0; \\ |\tau(\chi_0)| &= 1. \end{aligned} \tag{11b}$$

If we write by (11a),

$$\prod_{i=1}^{s} \chi_{k_i}(a^{-1}) = \chi(a);$$

$$\sum_{a \in K} \chi(ac_{s+1}^{-1})\zeta^{tr(a)} = \chi(c_{s+1}^{-1}) \sum_{a \in K} \chi(a)\zeta^{tr(a)} = \chi(c_{s+1}^{-1})\tau(\chi);\ c_{s+1} \neq 0,$$

and employ (11) and (11b) noting that there are $k$ characters $\chi_k$, and that $\chi_0$ is a $\chi_k$,

$$|T| \leqq \prod_{i=1}^{s} \sum_{\chi_{k_i}} |\chi_{k_i}(c_i^{-1})| \, ||\tau(\chi)|| \, \tau(\chi_{k_i})|$$

$$|T| \leqq \prod_{i=1}^{s} (1 + (k_i - 1)p^{n/2})p^{n/2}. \tag{12}$$

Let us now suppose that $c_{s+1} = 0$. Then we have $\zeta^{tr(ac_{s+1})} = 1$ in (11) and

$$\sum_{\substack{a \in K \\ a \neq 0}} \chi(c_i^{-1}a^{-1}) \leqq p^n - 1$$

and then

$$|T| \leqq (p^n - 1) \sum_{i=1}^{s} (1 + (k_i - 1)p^{n/2}) \tag{13}$$

Hence (10) gives for $c_{s+1} \neq 0$,

$$\left| N - \frac{(p^n - 1)^s}{p^n} \right| \leqq \prod_{i=1}^{s} (1 + (k_i - 1)\, p^{n/2})p^{-n/2}, \tag{14}$$

and for $c_{s+1} = 0$,

$$\left| N - \frac{(p^n - 1)^s}{p^n} \right| \leqq \frac{(p^n - 1)}{p^n} \prod_{i=1}^{s} (1 + (k_i - 1)\, p^{n/2}). \tag{15}$$

Whence we have

THEOREM II. *If N is the number of solutions of (1) under the condition (2), then for* $c_{s+1} \neq 0$, $(p^n - 1, a_i) = k_i$,

$$\frac{(p^n - 1)^s}{p^n} - D \leqq N \leqq \frac{(p^n - 1)^s}{p^n} + D, \tag{16}$$

*and for* $c_{s+1} = 0$,

$$\frac{(p^n - 1)^s}{p^n} - D_1 \leqq N \leqq \frac{(p^n - 1)^s}{p^n} + D_1, \tag{17}$$

*where*

$$D = p^{-n/2} \prod_{i=1}^{s} (1 + (k_i - 1)\, p^{n/2}),$$

*and*

$$D_1 = \frac{p^n - 1}{p^n} \prod_{i=1}^{s} (1 + (k_i - 1)\, p^{n/2}).$$

From the above theorem we infer immediately that

$$N = \frac{(p^n - 1)^s}{p^n} + 0(p^{\frac{n(s - v)}{2}}), \tag{18}$$

where $v = 1$ or $0$ according as $c_{s+1} \neq 0$ or $c_{s+1} = 0$.

If we consider (1) without the restriction $x_1, x_2 \ldots x_s \neq 0$, we may find limits for the number of solutions by considering one or more of the $c$'s in (1) as being zero in turn and employ for each of the resulting equations the limits (16) and (17) and adding the results we obtain inferior and superior limits of somewhat the same type as (16) and (17) and in particular we infer that if $N_1$ is said number then

$$N_1 = p^{n(s-1)} + 0(p^{n(s-v)/2}); \tag{19}$$

$v$ having the same meaning as in (18).

[1] *Vorlesungen über Zahlentheorie*, Leipzig, Hirzel, 1927, Bd. I, 83–87. Cf. also Dickson, *Modern Elementary Theory of Numbers*, University of Chicago Press, 1939, pp. 272–276.

[2] These PROCEEDINGS, **34**, 258–263 (1948). This article will be referred to as E. Characters defined by a general finite field have been considered by a number of writers. Cf. the paper referred to in our footnote 5 below and the references given therein as well as Davenport, *Acta Mathematica*, **71**, 99–121 (1939).

[3] This term is employed for the reason that the special $R$ consisting of the residue classes modulo $m$, our units correspond to the integers less than $m$ and prime to it.

[4] In this paper we are using the symbols 1 and 0 both for the unity and zero elements in the complex field and the unity and zero elements in $R$. This should not cause confusion if we keep in mind that $\chi(a)$ is always the complex field while $a$ is always in $R$.

[5] This was proved for $n < p$ in $E$, and the argument employed there may be extended to the case $n \geq p$. However, the result was previously known in general, in fact it follows from the results in Stickelberger, *Math. Annalen*, **37**, 321 (1890), § 1, paragraph 5, since any finite field may be represented by a complete set of residue classes with respect to a prime ideal modulus in an algebraic field.

[6] Davenport and Hasse, *J. f. Math. (Crelle)*, **172**, 151–174 (1935) found, for the case $s = 2$ a formula related to our (10). See their relation (6.2), p. 173, and the value of $N$ given at the top of p. 174. Their result will be discussed in a future paper by the authors.

---

# *BAYES SOLUTIONS OF SEQUENTIAL DECISION PROBLEMS*

BY A. WALD AND J. WOLFOWITZ

COLUMBIA UNIVERSITY

Communicated by P. A. Smith, December 21, 1948

The foundations of a general theory of sequential decision functions were described by one of the authors in a paper[1] whose terminology and general notation will be adopted in this communication. In the present note we continue in some respects the work initiated by the authors in their paper[2] on the sequential probability ratio test.

Let $\{X_i\}$ be a sequence of independent, identically distributed chance variables which are either discrete or absolutely continuous. In either case let $p(a\ F)$ denote the probability function or the probability density function, as the case may be, at the point $a$, where $F$ is the distribution function of $X$. Let $x$ represent the infinite sequence of observations $x_1, x_2, \ldots, d = D(x)$ be a decision function with elements in the space $D^*$, and $W(F, d)$ be the weight function. The regularity conditions which insure the existence of the integrals in (1), (2) and (3) will, for reasons of brevity, not be given in the present note. It is assumed that $\operatorname*{Min}_{d} W(\xi, d|)$, defined in (3) below, exists. Let $n(x, D)$ be the sample number function, and $\xi$ a probability measure on the space $\Omega$ of admissible distribution functions $F$. We shall assume a constant cost $c$ per observation. Without loss of generality we may take $c = 1$. The risk function $r(F, D)$ is defined by

$$r(F, D) = \int_M W(F, [D(x)])dF^*(x) + \int_M n(x, D)dF^*(x), \quad (1)$$

where $M$ is the space of sequences $x$, and $F^*$ is the probability measure induced on it by the distribution function $F$. Finally we define

$$r(\xi, D) = \int_\Omega r(F, D)d\xi \quad (2)$$

and

$$W(\xi, d) = \int_\Omega W(F, d)d\xi. \quad (3)$$

We shall say that the decision function $D_0$ is a Bayes solution relative to the *a priori* distribution $\xi$ if

$$r(\xi, D_0) \leq r(\xi, D) \quad (4)$$

for all admissible functions $D(x)$.

For any probability measure $\xi$ on $\Omega$, one of the following three conditions must hold:

(1) $$\operatorname*{Min}_{d} W(\xi, d) < r(\xi, D)$$

for any $D$ for which $n(x, D) \geq 1$

(2) $\operatorname*{Min}_{d} W(\xi, d) \leq r(\xi, D)$ for all $D$ for which $n(x, D) \geq 1$

with the equality sign valid for at least one $D$ with $n(x, D) \geq 1$

(3) There exists a $D$ with $n(x, D) \geq 1$ such that

$$\operatorname*{Min}_{d} W(\xi, d) > r(\xi, D).$$

We shall say that a probability measure $\xi$ is of type 1, 2 or 3, according as it satisfies condition 1, 2 or 3.

Let $\omega$ be a measurable subset of $\Omega$. Its *a posteriori* probability, given $X_i = x_i$, $i = 1, \ldots, m$, and given that $\xi$ is the *a priori* probability distribution, is

$$p_m(\omega|\xi, x_1, \ldots, x_m) = \frac{\int_\omega \prod_{i=1}^{m} p(x_i|F)d\xi}{\int_\Omega \prod_{i=1}^{m} p(x_i|F)d\xi}. \qquad (5)$$

Since this *a posteriori* probability is a probability measure on $\Omega$ it must be one of the three types defined above.

Theorem 1. *A necessary and sufficient condition for a decision function $d = D_0(x)$ to be a Bayes solution relative to a given a priori distribution $\xi_0$ is that the following three relations be fulfilled for any sample point $x$, except perhaps on a set whose probability measure is zero when $\xi_0$ is the a priori distribution on $\Omega$:*

(*a*) For any $m < n(x, D_0)$ the *a posteriori* distribution $p_m(\omega|\xi_0, x_1, \ldots, x_m)$ is either of type 2 or of type 3.

(*b*) For $m = n(x, D_0)$, the *a posteriori* distribution $p_m(\omega|\xi_0, x_1, \ldots, x_m)$ is either of type 1 or of type 2.

(*c*) For $m = n(x, D_0)$ we have

$$\operatorname*{Min}_{d} W(\xi(x_1, \ldots, x_m), d) = W(\xi(x_1, \ldots, x_m), D(x)), \qquad (6)$$

where $\xi(x_1, \ldots x_m)$ is an *a priori* distribution which is also the *a posteriori* distribution corresponding to $\xi_0, x_1, \ldots, x_m$.

A class of probability measures $\xi$ on $\Omega$ will be said to be convex if, for any two elements $\xi_1$ and $\xi_2$ in the class and for any positive value $\lambda < 1$, the probability measure $\xi = \lambda\xi_1 + (1 - \lambda)\xi_2$ is an element of the class. For any element $d_0$ of $D^*$ let $C(i, d_0)$ denote the class of probability measures $\xi$ of type $i (i = 1, 2, 3)$ for which

$$W(\xi, d_0) = \operatorname*{Min}_{d} W(\xi, d). \qquad (7)$$

Theorem 2. *Let $d$ be any element of $D^*$. The class $C(1, d)$ and the set theoretic sum $C(d)$ of $C(1, d)$ and $C(2, d)$ are both convex.*

We shall say that a set $L$ of probability measures $\xi$ on $\Omega$ is a linear manifold if, for any real value $\alpha$ and any two elements $\xi_1$ and $\xi_2$ of $L$, $\xi = \alpha\xi_1 + (1 - \alpha)\xi_2$ is an element of $L$ if it is a probability measure on $\Omega$. The linear manifold $L$ will be said to be tangent to $C(d)$ if (*a*) the intersection of $L$ and $C(2, d)$ is not empty (*b*) the intersection of $L$ and $C(1, d)$ is empty. For any decision function $D(x)$ and for any element $d$ of $D^*$, let $L(D, d)$ denote the linear manifold consisting of all $\xi$ which satisfy

$$W(\xi, d) = r(\xi, D). \qquad (8)$$

Theorem 3. *Let $\xi_0$ be an element of $C(2, d)$ and let $D_0(x)$ be a decision function which requires at least one observation and is such that $W(\xi_0, d) = r(\xi_0, D_0)$. Then the linear manifold $L(D_0, d)$ is tangent to $C(d)$.*

Detailed applications of the above results to the interesting special case when both $\Omega$ and $D^*$ contain a finite number of elements have been made by the authors and will be reported in detail elsewhere, together with proofs of the above theorems and other results. The convexity properties of the sets $C$ in this case were described by one of the authors (Wald) in his invited address given at the Berkeley, California, meeting of the Institute of Mathematical Statistics, June 22, 1948.

[1] Wald, Abraham, "Foundations of a General Theory of Sequential Decision Functions," *Econometrica*, 15 (4), 279–313 (October, 1947).

[2] Wald, A., and Wolfowitz, J., "Optimum Character of the Sequential Probability Ratio Test," *Ann. Math. Stat.*, *XIX* (3), 326–339 (September, 1948).

---

# *PHARMACOLOGICAL MODIFICATION OF RESISTANCE TO RADIATION—A PRELIMINARY REPORT*

BY JOHN B. GRAHAM AND RUTH M. GRAHAM

FROM THE VINCENT MEMORIAL LABORATORY, VINCENT MEMORIAL HOSPITAL (THE GYNECOLOGICAL SERVICE), MASSACHUSETTS GENERAL HOSPITAL, BOSTON, MASSACHUSETTS

Communicated by Eric G. Ball, December 30, 1948

It has always been assumed that an individual's response to ionizing radiation is an unchanging and unalterable characteristic. We believe this assumption is erroneous. The curability of cancer patients by roentgen or radium therapy has been ascribed to the radiosensitivity of the tumor itself. Repeated attempts to correlate the type of tumor and its sensitivity to radiation, for the most part have been inconclusive. It is our impression that the response of the normal tissue is at least as important as that of the tumor cell. In 1947, one of us[1] reported a study of 73 cases of carcinoma of the cervix treated by x-radiation and radium. These cases included all stages of disease and histologically the tumors were either squamous or adenocarcinoma. Vaginal smears were taken on the patients during or immediately following their course of treatment and examined for cytological evidence of radiation. Thirty-six showed marked radiation changes in both normal and malignant cells and thirty-seven showed little or no response. Eighty-five per cent of those patients whose smears showed radiation reaction in both *normal* and malignant cells were "cured." On the other hand, in that group whose smears showed little or no response, 97% died or had recurrent tumor. We feel that this is evidence to substantiate the importance of changes in the normal cells, since in the poor response group even though the malignant cells disappeared from the smear, if the *normal cells* showed no radiation response the outcome was poor. The important feature of this observation seems to be that the response to radiation as far as cure is concerned rests not only upon the character of the tumor cell but also depends upon an unknown factor in the non-neoplastic tissue as indicated by the observed changes in the normal cells. The mere disappearance of the tumor cells from the smear may not indicate whether or not the tumor is controlled, but the study of the changes in the normal cells appears to be a reliable index of the probability of a cure.

Both animals and man show great variation in their ability to withstand total body radiation.[2, 3] In animals, that diversity may be observed not only in mortality figures but in the cellular response as well. Swiss mice weighing 18–25 g. were subjected to varying amounts of total body radiation from 400 to 600 r at 200 kv. Vaginal smears were done on female mice and oral smears on males to determine their cellular response. Table 1 shows both radiation response and mortality figures in male and female mice treated with increasing amounts of radiation. There is a direct correlation between cellular response and mortality. Not all animals die which show radiation response but, more important, with only a rare exception, all mice survive which *do not* show radiation changes.

TABLE 1

MORTALITY RATE AND CELLULAR RESPONSE IN INCREASING DOSES OF TOTAL BODY RADIATION IN SWISS MICE

| AMOUNT OF X-RAY | SEX | NO. OF ANIMALS | PERCENTAGE MORTALITY AT 40 DAYS | PER CENT OF ANIMALS SHOWING CELLULAR RESPONSE* |
|---|---|---|---|---|
| 400 r | m | 230 | 21 | Not done† |
| | f | 220 | 28 | 25 |
| 500 r | m | 50 | 16 | Not done† |
| | f | 70 | 26 | 44 |
| 550 r | m | 128 | 50 | 88 |
| | f | 162 | 73 | 90 |
| 600 r | m | 35 | 100 | Not done |
| | f | 35 | 100 | Not done |

* Cytological study was not performed on every animal but a significant sample was made of each group.

† Males were not studied in the initial phase of the work because of technical difficulty with oral secretions.

This demonstrates that animals as well as man vary in their sensitivity to radiation and that this sensitivity may be gauged by a study of normal cell reaction. The importance of this observation has two possible clinical implications provided methods could be found to modify this native sensitivity. Treatment which would raise the resistance might prove of value in protecting individuals from total body radiation, while substances which lowered the resistance to radiation might prove useful in the treatment of patients with carcinoma.

Hektoen and others have reported that foreign protein[4, 5] and estrogens[6] administered parenterally 9–10 days before x-ray treatment diminish the effect of radiation. Desoxycorticosterone has also been listed as protective in this regard but no dose or time of administration is given.[7]

TABLE 2

MORTALITY RATE IN SWISS MICE TREATED BEFORE OR AFTER 400 R RADIATION (200 KV.)*

| | | TIME OF ADMINISTRATION | | | | |
|---|---|---|---|---|---|---|
| SUBSTANCE | DOSE | 10 DAYS BEFORE X-RAY | IMMEDIATELY AFTER X-RAY | SEX | NO. OF ANIMALS | MORTALITY AT 40 DAYS IN % |
| Controls (radiated) | ... | .. | .. | m | 230 | 21 |
| | | | | f | 220 | 28 |
| Horse serum (1) | 0.5 cc. | x | .. | m | 25 | 0 |
| | | | | f | 25 | 0 |
| | 1.0 cc. | .. | x | m | 32 | 60 |
| | | | | f | 16 | 50 |
| Testosterone (2) propionate in oil | 1.25 mg. | x | .. | m | 25 | 24 |
| | | | | f | 25 | 24 |
| | 1.25 mg. | .. | x | m | 25 | 52 |
| | | | | f | 25 | 68 |
| Estradiol benzoate in oil (2) | 0.083 mg. | x | .. | m | 30 | 6 |
| | | | | f | 20 | 10 |
| | 0.083 mg. | .. | x | m | 34 | 32 |
| | | | | f | 17 | 48 |
| Stilbestrol in oil (3) | 0.1 mg. | x | .. | m | 35 | 82 |
| | | | | f | 14 | 42 |
| | 0.1 mg. | .. | x | m | 25 | 8 |
| | | | | f | 24 | 33 |
| Desoxycorticosterone in oil | 0.1 mg. | x | .. | m | 25 | 16 |
| | | | | f | 25 | 20 |
| | 0.1 mg. | .. | x | m | 25 | 52 |
| | | | | f | 25 | 52 |

* Controls: 20 mice injected with each substance, not radiated. 2 mice dead at 40 days. Mort. 2%. 100 mice not injected, not radiated. 1 mouse dead at 40 days. Mort. 1%.

NOTES: (1) Supplied by Mass. State Antitoxin Laboratory. (2) Supplied by Schering Corp., Bloomfield, N. J. (3) Supplied by Burroughs & Wellcome, New York, N. Y.

We have administered a variety of steroid substances and foreign protein intraperitoneally to mice 10 days before or immediately after total body radiation. Table 2 shows mortality figures in male and female Swiss mice given 400 r total body radiation at 200 kv. and the modification of this mortality rate by the administration of foreign protein, testesterone proprionate, estradiol benzoate, stilbestrol and desoxycorticosterone. It is obvious that there is a distinct variation between the sexes, especially marked in the stilbestrol pretreated group. The time of administration is important. Different effects are encountered depending on whether the substance was injected 10 days before or immediately after radiation, e.g., the groups injected with horse serum. With the limited number of animals employed only the more striking differences are regarded as significant. For this reason in later experiments both total body weights of the animals and the cytological response have been included in our evaluation. This is shown in table 3 which shows comparable data in a small series of animals treated before radiation with two different substances, foreign protein and desoxycorticosterone. In these two experiments weights of animals and the cytological effect of radiation were taken into consideration as well as the mortality figures. In animals treated with desoxycorticosterone the death rate affords scant evidence of any change in sensitivity to radiation since an equal percentage of animals died as in the control radiated group. However, when both weights and cytological evidence of radiation are considered, the animals treated with desoxycorticosterone are found to have been extremely sensitive to the radiation administered and the protein group highly resistant. The radiated controls fall in the middle group, some animals being sensitive, some resistant. We believe that the cytological response may be a useful index in the study of the effects of radiation and that by its use finer differences can be observed. It has the additional advantage of affording study of cellular response without the sacrifice of an animal.

TABLE 3

EFFECT OF TREATMENT WITH HORSE SERUM (0.5 CC.) OR DESOXYCORTICOSTERONE (0.1 MG.) 10 DAYS BEFORE RADIATION (400 R) IN FEMALE SWISS MICE

| | PROTEIN | DESOXYCORTICOSTERONE | RADIATED CONTROLS |
|---|---|---|---|
| No. of mice | 12 | 12 | 8 |
| Mortality at 40 days | 0 | 20% | 20% |
| Initial body weight not regained at 20 days | 1 | 9 | 5 |
| Radiated response as judged by vaginal smear | 10 slight<br>2 marked | 11 marked<br>1 slight | 4 marked<br>4 slight |

From these data we conclude that an individual's sensitivity to ionizing radiation may be either enhanced or diminished by the administration of certain steroids or foreign protein before or after radiation. It is possible that this observation may have some bearing on survival from total body radiation and on the effectiveness of radiation treatment of cancer.

This work was performed under the auspices of the Office of Naval Research, U. S. Navy.

We are greatly indebted to Dr. Joe V. Meigs for his aid and encouragement in this endeavor. We are also indebted to Dr. Milford Schulz of

the Radiology Department of the Massachusetts General Hospital for the radiation of the experimental animals.

[1] Graham, R. M., *Surg. Gynec. & Obst.*, **84,** 153–173 (1947).

[2] Henshaw, P. S., *J. Nat. Cancer Inst.*, **4,** 485–501 (1944).

[3] Larkin, J. C., *Am. J. Roentgenol.*, **55,** 525–532 (1946).

[4] Hektoen, L., *J. Infect. Dis.*, **22,** 28–33 (1918).

[5] Auer, J., and Witherbee, W. D., *J. Exper. Med.*, **33,** 791–813 (1921).

[6] Treadwell, H., Gardner, W. V., and Laurence, J. H., *Endocrinology*, **32,** 161–164 (1943).

[7] Ellinger, F., *Radiology*, **50,** 234–243 (1948).

---

# PHYSICAL SYSTEMS OF CURVES IN SPACE*

By Edward Kasner and John De Cicco

Departments of Mathematics, Columbia University, New York, and Illinois Institute of Technology, Chicago

Communicated December 30, 1948

1. Consider a field of force whose force vector acting at any point, defined by the position vector $r = (x, y, z)$, of a given region of space, is $F = (\phi, \psi, \chi)$. The rectangular components $(\phi, \psi, \chi)$ depend only on the position of the point, and are assumed to be continuous and to have continuous partial derivatives of first and second orders over the given region of space. We omit the trivial case where the force vector is identically zero for which the trajectories are all straight lines.

There is no loss in generality in assuming that a particle traveling in this field is of unit mass. Dots denote total differentiation with respect to the time $t$ and the subscript $s$ denotes total differentiation with respect to the arc length. Subscripts $x$, $y$, $z$ signify partial differentiation with respect to the appropriate variable.

A *system* $S_k$ of $\infty^5$ curves in the given field of force consists of curves along which a constrained motion is possible such that the osculating plane at each point contains the force vector $F$, and the pressure $P$ is proportional to the normal component $N$ of $F$. Thus $P = kN$, where $k \neq -1$.

This is the extension of our previous two-dimensional theory of physical systems of curves, outlined in papers published in these PROCEEDINGS during the past two years.

*The vector differential equation of a system $S_k$ of $\infty^5$ space curves is*

$$\ddot{r} = (1 + k)\, F - \frac{k(F\dot{r})}{\dot{r}^2}\, \ddot{r}. \qquad (1)$$

The important special cases of physical interest of a system $S_k$ are

(*a*) dynamical trajectories, $k = 0$;
(*b*) general catenaries, $k = 1$;
(*c*) generalized brachistochrones, $k = -2$;
(*d*) velocity curves, $k = \infty$.

2. *Property I.* The osculating planes of all the $\infty^3$ curves of a system $S_k$ which pass through a fixed point, form a pencil of planes with axis in the direction of the force vector acting at the point.

By varying $k$, it is found that *all the $\infty^2$ curves of all possible systems $S_k$ which pass through a given lineal element E, possess the same osculating plane and also the same torsion at E.*

The curves of all possible systems $S_k$ which pass through a given point and have the same torsion there are $\infty^3$ in number. Their tangent lines through the point generate a quadric cone. By varying the torsion, we obtain a pencil of quadric cones, all of which contain the force vector $F$ acting at the given point, and have the same tangent plane along $F$. This is the osculating plane of the line of force at the given point.

3. *The centers of spherical curvature corresponding to the points of the curves of a system $S_k$ are given by the vector equation*

$$S = r + \frac{v^2(F_s \times r_s) - (3 + k)(Fr_s)(F \times r_s)}{(1 + k)\ (r_s F F_s)}, \qquad (2)$$

*where v denotes the speed.*

*Property II.* The osculating spheres of the $\infty^1$ curves of a system $S_k$ passing through a given point in a given direction form a pencil. *Their centers describe the straight line*

$$r_s \cdot (S - r) = 0, \qquad (1 + k)F_s \cdot (S - r) - (3 + k)(Fr_s) = 0. \quad (3)$$

By varying $k$, *the straight lines for all systems $S_k$ that correspond by Property II to the lineal element E, form a parallel pencil in the plane normal to E.*

4. Hold the point $r = (x, y, z)$ fixed and vary the direction $r_s$. The straight lines (3) of Property II form a congruence $\Omega$. If $G$ denotes the vector

$$G = [F_x \cdot (S - r),\ F_y \cdot (S - r),\ F_z \cdot (S - r)], \qquad (4)$$

the congruence $\Omega$ may be written in the form

$$(S - r) . r_s = 0, \qquad [(3 + k)F - (1 + k)G] . r_s = 0. \qquad (5)$$

*Property III. The congruence $\Omega$ is composed of the secants of a twisted cubic curve $\Gamma$.* This cubic $\Gamma$ passes through the fixed point in the direction of the force vector acting at the given point.

The vector equation of $\Gamma$ is

$$(3 + k)F - (1 + k)G = (1 + k)\lambda\ (S - r), \qquad (6)$$

*where $\lambda$ is an arbitrary scalar.*

5. *The locus of the centers of spherical curvature of the $\infty^2$ curves of a system $S_k$ touching a given plane at a given point is a quadric surface passing through the twisted cubic curve $\Gamma$.*

If $\pi$ is a vector normal to the given plane, the quadric surface is

$$(\pi, S - r, (1 + k)G - (3 + k)F) = 0. \qquad (7)$$

*This quadric surface (7) is a rectangular hyperboloid if and only if the corresponding fixed plane contains the curl $F = (\chi_y - \psi_z, \phi_z - \chi_x, \psi_x - \phi_y)$ of the force vector $F = (\phi, \psi, \chi)$ acting at the fixed point.*

A field of force is *conservative* if and only if every quadric surface (7) is a rectangular hyperboloid; or if and only if the three asymptotes of the twisted cubic curve $\Gamma$ are mutually orthogonal.

6. The straight line which corresponds according to Property II, to a lineal element $E$ belonging to a line of force is parallel to the binormal of the line of force. Its perpendicular distance from the given point is $(3 + k)/(1 + k)$ times the radius of curvature of the line of force.

This is related to the discussion of rest trajectories and the theorem about the ratio of 1:3 of curvatures.

---

* Presented to the American Mathematical Society, 1949.

[1] Kasner, "Differential-Geometric Aspects of Dynamics," Princeton Colloquium Lectures, *Am. Math. Soc. Publ.*, (1913, 1934, 1948).

[2] Kasner, "Dynamical Trajectories: The Motion of a Particle in an Arbitrary Field of Force," *Trans. Am. Math. Soc.*, **8**, 135–158 (1907). See also a series of papers in *Trans. Am. Math. Soc.*, **7–11** (1906–1910).

[3] Kasner, "Physical Curves," *Proc. Nat. Acad. Sci.*, **34**, 68 (1948).

[4] Kasner and De Cicco, "Transformation Theory of Physical Curves," *Ibid.*, **34**, 169 (1948).

[5] De Cicco, "Constrained Motion upon a Surface under a Generalized Field of Force," *Bull. Am. Math. Soc.*, **53**, 993–1001 (1947).

[6] Kasner and Mittleman, "A General Theorem on the Initial Curvatures of Dynamical Trajectories," *Proc. Nat. Acad. Sci.*, **28**, 48–52 (1942).

---

# *CONJUGATION BETWEEN DOUBLE MONSTERS AND SINGLE ANIMALS IN PARAMECIUM BURSARIA**

By Tze-Tuan Chen

Department of Zoology, University of California, Los Angeles

Communicated by T. M. Sonneborn, December 11, 1948

In the cultures of a Canadian clone (Can18) of *Paramecium bursaria*, double monsters, chains of two animals, occur frequently. When one animal has divided, the two daughter animals fail to separate. These two daughter animals then develop into the adult form, still connected by a slender protoplasmic bridge. Typically, each of the two components of the monster has a micronucleus and a macronucleus. In the present work these double monsters were isolated at various times and mated with the single animals of a U. S. clone (McD6).[1] Nuclear phenomena during such conjugations were studied.

Under proper conditions, when single animals of these two clones are mixed, the agglutinative mating reaction occurs immediately and pairs are soon formed. This is also true when double monsters of the Canadian clone are mixed with single animals of the U. S. clone. A double monster may conjugate with one or two (occasionally more) single animals. In most cases, one single animal is attached to the anterior component of the double monster and at the adoral surface. Nuclear changes occur in all conjugants whether the double monster conjugates with one or more single animals.

In each conjugant group of the most common type, consisting of a double monster and a single animal attached to the anterior half of the double monster, three micronuclei are present. The single animal contains one, the double monster has two—one in the anterior component, the other in the posterior component. Nuclear changes occur in all three micronuclei. The posterior component of the double monster does not come in contact with the single animal, yet the micronucleus in this half undergoes the usual nuclear changes. This indicates that the influence of the contact between the anterior component and the single animal must have passed to the posterior component through the slender protoplasmic bridge.

The nuclear changes occur at about the same rate as those in ordinary conjugating pairs. The conjugants separate at about the same time as they would in ordinary conjugation. The nuclear changes among the three micronuclei in the conjugant group are usually synchronous. In the following paragraphs these nuclear changes and the time relationships of stages (at 27°C.) will be described.

Each micronucleus undergoes three pregamic divisions which are identical with those occurring in ordinary conjugating pairs.[2] The first pregamic division is a long process involving complicated prophase changes and requiring approximately 23 hours for completion. During the first 10 hours the spindle-shaped micronucleus gradually swells up and becomes nearly spherical. The transformation of the micronucleus to a crescent begins approximately 12 hours after the onset of conjugation. By the sixteenth hour, these crescents have become greatly enlarged and their chromatin diffused. By the twentieth hour, the nucleus has rounded up, becoming ovoid or spherical, and individual chromosomes are visible. By the twenty-second hour, late prophase or metaphase is reached. By the twenty-third hour, in most conjugant groups, the anaphase or telophase stage is reached; in some, daughter nuclei are formed. Daughter nuclei are similar in size and structure. One degenerates soon after its formation, while the other remains to undergo the second pregamic division.

The second pregamic division occurs 24 or 25 hours after onset of conjugation and is completed in one or two hours. The two resulting nuclei are also similar in size and structure. One of them soon degenerates, and the remaining one undergoes the third pregamic division.

The third pregamic division is also consummated in a short time and ends approximately 28 hours after the onset of conjugation. Of the two resulting pronuclei, one is migratory, the other stationary. These two pronuclei may also differ in shape. This differentiation of the two pronuclei takes place in all three individuals of the conjugant group, even though the posterior half of the double monster does not mate with any animal. Exchange of pronuclei takes place only between the anterior half of the double monster and the single animal, the exchange occurs soon after the formation of the pronuclei, 28 to 29 hours after the onset of conjugation. But the migratory pronucleus in the posterior half of the double monster moves to the vicinity of the mouth region of its cell as if a conjugant were attached to that part of its body. (In some conjugant groups a bulge appears at this region.) This fact seems to indicate that the path of the migratory pronucleus is predetermined, and confirms my earlier finding in conjugation of three animals.[3]

No case has been observed in which the migratory pronucleus in the posterior half of the double monster breaks through the cell membrane; instead, it fuses with the stationary pronucleus in the same half to form a synkaryon. Autogamy thus occurs in the posterior half of the double monster.

Three divisions occur in all three synkarya before the conjugants separate, as in the case of ordinary conjugating pairs. The first division usually occurs 29 to 31 hours after the onset of conjugation. One of the two nuclei resulting from this division degenerates. Two further nuclear divisions occur at 33 to 36 hours. By the thirty-sixth hour, ex-conjugants may be found.

When the conjugants separate, the ex-conjugant double animal usually contains eight nuclei, while the single animal contains four nuclei. Some of these nuclei develop into macronuclear anlagen, others become the micronuclei. The number of nuclei that become macronuclear anlagen is variable. In most of the ex-conjugant single animals, two (occasionally three) of the four nuclei become anlagen. In the ex-conjugant double animals, two or four or five of the eight nuclei may develop into anlagen.

Variation in the number of nuclei developing into anlagen has also been found in ordinary conjugation in this species of Paramecium.

* This work was aided by grants from the Committee for Research in Problems of Sex, the NATIONAL RESEARCH COUNCIL; from the Joseph Henry Fund of the NATIONAL

ACADEMY OF SCIENCES; and from the Permanent Science Fund of the American Academy of Arts and Sciences.

[1] Both the Canadian and U. S. clones belong to Variety II. The former belongs to mating type J; the latter to type K.

[2] Chen, T. T., *J. Hered.*, **31**, 185–196 (1940); *J. Morph.*, **79**, 125–261 (1946). Wichterman, R., *Turtox News*, **26**, 2–10 (1948).

[3] Chen, T. T., *J. Morph.*, **78**, 353–395 (1946).

---

# *THE NATURE OF THE SPECIFICITY OF METHYL GREEN FOR CHROMATIN*

BY ARTHUR W. POLLISTER AND CECILIE LEUCHTENBERGER

DEPARTMENT OF ZOOLOGY, COLUMBIA UNIVERSITY

Communicated by L. C. Dunn, December 11, 1948

Of all substances which bulk large in the chemical composition of cellular structures, desoxypentose nucleic acid is perhaps the most clearly demonstrable by application of cytochemical methods to cells. For each one of the three components of nucleotide residues there is a specific method which enables the cytologist to visualize it on slides: the purines and pyrimidines have natural specific absorption in the ultra-violet spectrum; the Feulgen reaction reveals the desoxypentose; and the staining with basic dyes is an indication of the orthophosphoric acid component. It is this last specificity which many histologists might wish to question. It has been generally realized, since the chemical properties of nucleic acid became known, that for the most part the usual binding of basic dye in histological staining of the nucleus must be due to formation of a dye-nucleate salt (see especially Michaelis[1, 2]). However, to what extent the intensity of basic staining by any particular method can be strictly interpreted to indicate relative amounts of nucleic acid has not been so clear. It has been shown again and again that under the conditions of standard staining procedures the basic dyes color many structures that do not contain nucleic acid, and that the intensity of the color in chromatin is subject to many influences from both

TABLE 1

METHYL GREEN STAINING OF CYTOPLASM AND NUCLEI OF EPITHELIAL CELLS OF MOUSE LIVER

| NO. OF EXPERIMENT | PART OF CELL MEASURED | STAIN | MEAN EXTINCTION |
|---|---|---|---|
| 1 | Cytoplasm | Methyl green | 0.014 ± 0.001 |
| 1 | Nucleus | Methyl green | 0.294 ± 0.012 |
| 2 | Nucleus | Methyl green | 0.265 ± 0.014 |
| 3 | Nucleus | Methyl green, Pyronin | 0.274 ± 0.007 |

*within and without the tissue. In such a situation the only sure way has been that of checking the amount of dye-nucleate staining by use of* specific nucleases, as Brachet[3] has done. In effect, this makes the problem one that can be attacked best by methods of quantitative chemical cytology. The present paper is concerned with demonstration, by photometric cytological analysis, that one basic dye, methyl green, can be used in such a manner that the staining is highly specific for desoxypentose nucleic acid, that it is quantitatively related to the amount of nucleic acid present, and that it is dependent upon the physical state of the nucleic acid.

Most common histological staining methods do not appear to be as highly specific for nucleic acid as has been suggested, for example, by Pollister and Ris.[4] With such dyes as Pyronin B, Pyronin Y or Basic Fuchsin we have found that there is a considerable residue of stainable substance in nuclei and cytoplasm after all nucleic acids and nucleotides have been removed chemically, table 2. These dyes also stain heavily such material as thyroid colloid, which, as Caspersson and Gersh[5] have shown photometrically, almost certainly contains no nucleic acid. However, as was originally pointed out by Unna,[6] methyl green seems to be an exception in that it stains nothing but chromatin.

TABLE 2

METHYL GREEN AND PYRONIN STAIN BEFORE AND AFTER REMOVAL OF NUCLEIC ACIDS WITH HOT TRICHLORACETIC ACID

| METHOD, MATERIAL, WAVE-LENGTH MEASURED | BEFORE TRICHLORACETIC ACID | AFTER TRICHLORACETIC ACID |
|---|---|---|
| Methyl green, mouse liver, $E_{625}$ | 0.296 ± 0.012 | 0.010 ± 0.002 |
| Pyronin, mouse liver, $E_{540}$ | 0.480 ± 0.014 | 0.110 ± 0.004 |

The method of staining which we have used closely follows that of Unna and Pappenheim.[7] The dye is a National Aniline product, C.I. No. 685, Certification No. NG 25, Total Dye Content 93%. Under this Color Index Number Conn[8] discusses a dye called ethyl green (synonym, methyl green) which he says is the "zinc chloride double salt of ethylhexamethylpararosanilin chlorobromide," molecular weight 653. As received, the dye contains a small amount of a violet compound which must be removed, for it does not have the specific properties of the green dye. This purification may be carried out just before use by Unna's method of shaking the dye with chloroform, in which the purple component alone is soluble. The purification does not seem to be necessary with Grübler's Iodine Green, presumably a closely related dye, which has the same staining properties as methyl green. For use 75 mg. of the purified dye was dissolved in 50 $cm.^3$ of the phenol-alcohol-glycerin solution of Unna-Pappenheim. The pH of this solution is about 6.3. The absorption maximum of the dye as determined in the Beckmann Spectrophotometer is about 630 mμ. Paraffin sections of vertebrate tissues, fixed in Carnoy's acetic alcohol (1:3), were stained in freshly purified, freshly prepared dye solution for 45 minutes at 56°C., blotted, then decolorized overnight in tertiary butyl alcohol (or in 95% ethyl alcohol, when methyl green was used without pyronin). This approximates a procedure which Michaelis,[2] has suggested as most likely to leave in the tissue only chemically bound dye. The intensity of stain in the cells was estimated by the absorption methods described by Pollister and Ris,[4] measuring the transmission of red light isolated from a tungsten lamp by a Farrand Interference Filter with a peak transmission at 630 mμ. The lower extinction values are from partial liver nuclei in 5 μ sections, the higher from whole nuclei in 8 μ sections. The whole nuclear diameter was projected onto the photosensory surface. The extinction values are means of a number of measurements, each mean being from a series of nuclei of approximately equal diameter. Unstained slides show an absorption equivalent to an extinction of 0.010 to 0.030; consequently values in this range indicate that there has been no staining reaction.

When slides have been stained as described above it is apparent at a glance that the methyl green stains only chromatin; the amount of color in either nucleolus or cytoplasm (table 1) being negligible. For any one type of nucleus the amount of dye bound is accurately reproducible, tables 1 and 2. The mean values for 5 μ sections of nuclei in four different experiments ranged from 0.265 to 0.296, the differences not being statistically significant. The amount of methyl green staining of chromatin is independent of the presence of another dye such as pyronin B (table 1) or basic fuchsin.

A series of experiments demonstrates what chemical component of the chromatin is responsible for this specific binding of methyl green. The methyl green is staining the nucleic acid of chromatin, not the protein, for after nucleic acid has been removed from the sections by 15 minutes' treatment with hot 5% trichloracetic acid (Schneider,[9] Pollister and Ris[4]) the chromatin no longer stains to any appreciable extent with methyl green, table 2. The methyl green is bound to desoxypentose nucleic acid, not to the pentose type; for two hours' action of ribonuclease on the sections has no effect upon the methyl green staining.' The mean extinction of ten untreated whole nuclei was 0.374 ± 0.020 that of ten after enzyme action was 0.395 ± 0.020, which is not statistically different. In short, the series of experiments demonstrates conclusively that, in material which has been fixed in acetic alcohol, *purified methyl green is highly specific for desoxypentose nucleic acid.*

In reproducibility and specificity the methyl green staining is similar to a properly controlled Feulgen nucleal reaction (Stowell[10], Di Stefano[11]), which suggests that the methyl green binding likewise is quantitatively related to the amount of desoxypentose nucleic acid in a nucleus. This is further supported by the fact that in diverse normal nuclei, with varying amounts of chromatin, the ratio of nucleal-regenerated fuchsin to methyl green staining is approximately the same, the ratio of extinctions being not far from 1.0. It seems likely that in the case of methyl green we are dealing with a chromatin staining which is solely chemical combination of dye and nucleic acid, presumably to form a methyl green nucleate salt.

Among other technical discoveries Unna noted that the methyl green staining of chromatin could be markedly reduced by pretreatment of sections with boiling water. Photometric measurements demonstrate this effect in a striking manner. When sections which have been treated with hot water are stained, the methyl green binding does not occur to any appreciable extent, table 3. Indeed the methyl green extinction after hot water is as weak as on sections from which the nucleic acid has been entirely removed, table 3; but the explanation of the loss of staining after hot water must be quite different, since it can be demonstrated that the treatment has not removed any of the desoxypentose nucleic acid from the chromatin (table 3) for: (*a*) the nucleal reaction is equally intense before and after hot water, showing that the desoxypentose content is unchanged; (*b*) there is no loss of purine or pyrimidine bases from the chromatin, for the 2537 Å. absorption is not reduced; and (*c*) gross analyses of liver before and after hot water, which Dr. Gerhardt Schmidt was kind enough to make for us, show conclusively that the treatment causes no reduction of the phosphoric acid, the third component of the desoxypentose nucleic acid. This last is most interesting for it shows that the part of the molecule which is presumably most closely related to the binding of the basic dye to form a methyl green nucleate salt is still present after hot water in as great an amount as when the chromatin stained intensely with methyl green. The obvious conclusion from these data is that the loss of methyl green staining after hot water treatment is a consequence of a physical change in the nucleic acid molecule. From results reported by Kurnick,[11] we can infer that this change is in the nature of a depolymerization. In test-tube experiments Kurnick found that while methyl green combined readily with highly polymerized desoxypentose nucleic acid, the dye would no longer combine with nucleic acid which had been depolymerized by the enzyme. Pending the results of experiments with the

TABLE 3

| METHOD, MATERIAL, WAVE-LENGTH MEASURED | BEFORE HOT WATER, MEAN EXTINCTION | AFTER HOT WATER, MEAN EXTINCTION |
|---|---|---|
| Methyl green, mouse liver, nuclei, $E_{625}$ | 0.340 ± 0.010 | 0.032 ± 0.004 |
| Feulgen reaction, mouse liver, nuclei, $E_{560}$ | 0.427 ± 0.014 | 0.419 ± 0.015 |
| Gross analysis (Schmidt), mouse liver, DNA phosphorus per 100 mg. of nitrogen | 1.43 mg. | 1.44 mg. |
| U. V. absorption, cartilage, *Amblystoma*, $E_{253}$, | 0.400 | 0.522 |

effect of purified thymonucleodepolymerase on tissue sections, it may be tentatively concluded that in cytological preparations also the methyl green staining of chromatin is dependent upon the presence of desoxypentose nucleic acid in a highly polymerized state, and that the hot water treatment causes depolymerization. If this is actually the effect of the hot water, it has not proceeded so far as to degrade the nucleic acid to lower acid-soluble polymers, for sections which have been treated first with hot water and then subjected to cold trichloracetic acid (after Schneider[9]) give as intense a Feulgen reaction as do untreated controls.

The more closely the methyl green reaction is examined the more conspicuous do its differences from other types of basic staining appear, and one feels impelled to look for special chemical properties of methyl green as an explanation. An obvious difference is that unlike all other common histological stains the methyl green colored ion carries two positive charges, which surely might be expected to give it unique properties as a stain. The divalent methyl green ion should have a much greater affinity for acidic groups like phosphoric acid than should the monovalent ions of other dyes and *a priori* one should therefore expect that methyl green, alone, would stain very strongly all acidic structures, and, when mixed with pyronin, it should be much the stronger staining component for acidic material. In the staining which has been described above this expected result is realized only to a limited extent. Where a structure is stainable with methyl green, as chromatin is, then the divalent cation completely predominates over the monovalent—that is, the pyronin exerts no measurable restriction on the methyl green reaction. Superimposed upon this readily understandable relationship between the two dyes, however, is another influence which appears as a sharp restriction of the methyl green reaction to one particular acidic component of the cell, the phosphoric acid of the desoxypentose nucleic acid. It seems possible that this also may be explained by the structure, a biologist is tempted to say the "morphology," of the molecule of methyl green. Michaelis[2] has reported that the spectroscopic characteristics of basic dye bound in tissue staining show conspicuous differences from those of the dye-nucleate in test tubes, which are best interpreted as evidence of a specific "fit" or structural relationship between the molecules of the dye and those of the nucleic acid. It is not difficult to imagine that the high specificity of methyl green for highly polymerized desoxypentose nucleic acid may be explained in somewhat the same way.

In the normal tissues the Feulgen-methyl green ratio is fairly constant; in other words, there appears to be a definite proportion of the total desoxypentose nucleic acid which is in the physical state suitable for combination with methyl green. Any decrease in relative amount of this presumably highly polymerized nucleic acid should lead to decrease in the methyl green reaction and should increase the Feulgen-methyl green ratio, above the normal range about a mean value of 1.0. Thus, if it is combined with a parallel study of slides on which the Feulgen reaction has been performed, the methyl green staining is potentially an indicator of a physical alteration of the desoxypentose nucleic acid molecule. Such changes may be expected under many cytological conditions, especially, for example, where chromatin is degenerating. We have recently been examining with these techniques the pycnotic degeneration that occurs in cells of a tumor implanted into a new host (as originally described by Rössle[12]). It is interesting that the first major change in the chromatin is a decrease in the intensity of the methyl green reaction, leading to such Feulgen-methyl green ratios as 3 : 1 or eventually perhaps as high as 8 : 1.

[1] Michaelis, L., *Einfuehrung in die Farbchemie fuer Histologen*, 1910.

[2] Michaelis, L., *Cold Spring Harbor Symposia*, XII, 131–142 (1947).

[3] Brachet, J., *Embryologie Chimique*, Masson et Cie., 1944.

[4] Pollister, A. W., and Ris, H., *Cold Spring Harbor Symposia*, XII, 147–157 (1947).

[5] Caspersson, T., and Gersh, I., *Anat. Rec.*, **78**, 303–321 (1940).

[6] Unna, P. G., *Histochemie der Haut*, 1928.

[7] Unna, P. G., *Enzyk. Mik. Tech.*, (1910).

[8] Conn, H. J., *Biological Stains*, (1946).

[9] Schneider, W. C., *J. Biol. Chem.*, **161**, 293–303 (1945).

[10] Stowell, R. E., *Stain Tech.*, **20**, 45–48 (1945).

[11] Di Stefano, H. S., PROC. NAT. ACAD. SCI., **34**, 75–80 (1948).

[12] Rössle, R., *Ueber die Anfaenge der krebsigen Neubildung bei Impfgeschwuelsten*, (1936).

---

## ERRATUM

In Section 4 of our paper "Some Properties of Rotational Flow of a Perfect Gas," March, 1948 (concerning Massotti Flow) the statement near the middle of the Section, that the vorticity equation

$$\operatorname{curl}\left[\frac{\mathfrak{w} \times \operatorname{curl} \mathfrak{w}}{1 - w^2}\right] = 0$$

is automatically satisfied, is incorrect and invalidates the remainder of the Section.

Actually, it is easily seen that this equation introduces the additional restriction that

$$1 - \left(\frac{\partial \phi}{\partial x}\right)^2 - \left(\frac{\partial \phi}{\partial y}\right)^2 = f(s)g(x, y)$$

where $f$ and $g$ are arbitrary functions. This restriction limits the solutions known at present to those for which $w^2$ = const.

P. NEMÉNYI AND R. PRIM

PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES

Volume 35 March 15, 1949 Number 3

# *NATIONAL ACADEMY OF SCIENCES (DISTRIBUTION AND AGES OF MEMBERSHIP)*

By Fred E. Wright, *Home Secretary*

Communicated January 29, 1949

The National Academy of Sciences was established by an Act of Congress, approved on March 3, 1863 by Abraham Lincoln. The number of charter members was 48; of these two declined membership. In the Act, the membership was limited to 50; the Academy was given power "to make its own organization including constitution and bylaws"; "to fill all vacancies created by death or resignation"; "to provide for the election of foreign and domestic members, the division into classes," "and to report the same to Congress." In addition, "the Academy shall, whenever called upon by any department of the Government, investigate, examine, experiment and report upon any subject of science or art, the actual expense of such investigations, examinations, experiments and reports to be paid from appropriations which may be made for the purpose, but the Academy shall receive no compensation whatever for any services to the Government of the United States."

The primary function of the Academy is therefore to serve as scientific adviser to the Government on problems in science and applied science for the solution of which a government department may seek unbiased, expert aid. Election to membership in the Academy means recognition of unusual ability; and it carries with it an obligation toward the Government. As a result, Academy membership is restricted largely to scientists who through research have contributed to increase in knowledge and are in position, upon request, to serve the Government in the solution of its problems in science.

The total Academy membership on July 1, 1948, numbered 435, including 5 members emeriti. Study of the membership with reference to distribution within the United States, to age and to fields of activity in science has been made at irregular intervals in the past. Early studies were reported upon by Raymond Pearl [Proc. Nat. Acad. Sci., **11**, 752–768 (1925)]. In 1933 and in 1935 F. E. Wright, as Home Secretary, continued the studies and circulated the results in mimeograph form to the memership of the Acad-

TABLE 1

DISTRIBUTION OF MEMBERSHIP BY STATES AND COUNTRIES (JULY 1, 1948)

| | NO. BORN | NO. NOW |
|---|---|---|
| New England: | | |
| Maine | 5 | 2 |
| Vermont | 2 | 0 |
| New Hampshire | 4 | 0 |
| Massachusetts | 35 | 68 |
| Connecticut | 6 | 20 |
| Rhode Island | 1 | 2 |
| Total | 53 | 92 |
| North Atlantic States: | | |
| New York | 42 | 75 |
| New Jersey | 4 | 28 |
| Pennsylvania | 24 | 16 |
| Maryland | 8 | 17 |
| District of Columbia | 2 | 29 |
| Delaware | 3 | 1 |
| Total | 83 | 166 |
| South Atlantic States: | | |
| Virginia | 7 | 4 |
| West Virginia | 2 | 0 |
| North Carolina | 2 | 2 |
| South Carolina | 2 | 0 |
| Georgia | 4 | 0 |
| Florida | 0 | 2 |
| Total | 17 | 8 |
| North Central States: | | |
| Michigan | 12 | 6 |
| Wisconsin | 20 | 13 |
| Minnesota | 5 | 4 |
| North Dakota | 1 | 0 |
| South Dakota | 4 | 0 |
| Ohio | 25 | 0 |
| Indiana | 18 | 4 |
| Illinois | 32 | 35 |
| Iowa | 12 | 4 |
| Nebraska | 5 | 0 |
| Missouri | 12 | 12 |
| Kansas | 10 | 0 |
| Total | 156 | 78 |

| | NO. BORN | NO. NOW |
|---|---|---|
| South Central States: | | |
| Kentucky | 5 | 0 |
| Tennessee | 4 | 1 |
| Arkansas | 0 | 0 |
| Oklahoma | 2 | 0 |
| Alabama | 1 | 0 |
| Mississippi | 0 | 0 |
| Louisiana | 1 | 0 |
| Texas | 4 | 6 |
| Total | 17 | 7 |
| Rocky Mountain States: | | |
| Montana | 0 | 0 |
| Wyoming | 1 | 0 |
| Idaho | 0 | 0 |
| Colorado | 5 | 1 |
| Utah | 3 | 1 |
| Nevada | 0 | 0 |
| New Mexico | 1 | 1 |
| Arizona | 0 | 1 |
| Total | 10 | 4 |
| Pacific States: | | |
| Washington | 3 | 0 |
| Oregon | 1 | 0 |
| California | 17 | 80 |
| Total | 21 | 80 |

Foreign Born:

| | | | |
|---|---|---|---|
| Austria | 4 | Italy | 1 |
| Canada | 11 | Jugoslavia | 1 |
| Denmark | 2 | Mexico | 1* |
| England | 5 | Norway | 2 |
| Estonia | 1 | Poland | 3 |
| France | 1 | Russia | 9 |
| Germany | 11 | Scotland | 3 |
| Holland | 4 | Sweden | 6 |
| Hungary | 4 | Switzerland | 4 |
| India | 2* | Syria | 1* |
| Ireland | 1 | Ukraine | 1 |

* American parents.

emy. In 1945 Edwin B. Wilson published a short report on "Vital Statistics of the National Academy of Sciences" [PROC. NAT. ACAD. SCI., **31**, 200–202 (1945)] and stressed the slowly rising average age of members at election.

Request has been made recently for a revision of the 1935 statement, based on the present membership of the Academy. Since 1935 the number of Academy members has increased from 289 to 435; this rapid growth has been chiefly due to action by the Academy in 1942 which raised the limit from 15 to 30 in the number of men who may be elected at any annual meeting. The result has been that more than one-half the Academy members, now living, were elected during the ten-year period 1939 to 1948, inclusive.

*Regional Distribution of Members.*—In table 1 data on the distribution of the membership in the several states are given. The table shows that in New England, the North Atlantic States and California the number of Academy resident members greatly exceeds the number of Academy members born therein; the ratio is 337 to 157. In the North Central States, 156 members were born, but only 78 are now resident therein; in the South Atlantic, South Central and Rocky Mountain States, 44 members were born, but only 19 now live there. In 17 states no Academy member resides; in these states, however, 73 members were born. In seven states no member was born; in five of these no member now resides. The trend of Academy members is naturally toward the centers of learning and of active research in science.

Seventy-eight Academy members were born in foreign countries; four of these are of American parentage. Twenty-two are from Canada, England, Scotland, Ireland and India; 10 from Sweden, Norway and Denmark; 23 from Germany, Austria, Holland and Switzerland; 16 from Russia, Poland and Hungary. The number of Academy members whose native tongue is not English is 54, or 12.4 per cent of the total membership.

The data on distribution of the Academy membership in universities and research organizations, including the Government, are listed in table 2. Of the total membership, 315 (72.4 per cent) are connected with universities and colleges; 25 (5.7 per cent) serve in Government positions, while 62 (14.3 per cent) are connected with endowed research groups; and 33 (7.6 per cent) with state and industrial research laboratories.

*Age Distribution of Membership.*—Table 3 presents the numbers of members in each age group and in the various Academy sections at the time of election and on July 1, 1948; also the average age and median age of the members at election and on July 1, 1948. In the preparation of this table the age of each member, in years and tenths of a year, at the time of his election and on July 1, 1948, was first ascertained. To facilitate the computations the average date (April 27) of the annual Academy meeting, at which elections are held, was used. Various methods may be used for computing the age of a person to tenths of a year of 365 or 366 days and the results obtained vary slightly with the method employed; but they are not significant in this case and may be neglected. The method adopted in this paper for computing the ages of members to tenths of a year may best be illustrated by an example; if the birthday of a member falls between April 27 and June 1 his age at election is, say, 41.9 years; whereas if it occurs between March 21 and April 26 of the same year his age is 42.0 years. With the aid of a short table prepared on this basis it is a simple task to ascertain the age of a member to tenths of a year. The fact that Pearl in his article of 1925 used April 15 as the date of election of each member was recognized by him to be unimportant in his study of vital statistics of Academy members.

TABLE 2

DISTRIBUTION OF MEMBERSHIP IN UNIVERSITIES AND RESEARCH ORGANIZATIONS (JULY 1, 1948)

*UNIVERSITIES*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Harvard | 51 | Wisconsin | 12 | Iowa | 4 | Smith | 1 |
| California | 34 | Princeton | 10 | Virginia | 4 | Haverford | 1 |
| Chicago | 27 | Cornell | 9 | Indiana | 4 | Brooklyn | 1 |
| Columbia | 25 | Pennsylvania | 8 | Rochester | 3 | Missouri | 1 |
| Yale | 18 | Wash. U., St. Louis | 8 | Brown | 2 | Swarthmore | 1 |
| Cal. Tech. | 17 | Illinois | 7 | North Carolina | 2 | Rice Inst. | 1 |
| Stanford | 17 | Texas | 5 | U. of St. Louis | 2 | Tennessee | 1 |
| M.I.T. | 13 | Michigan | 4 | Northwestern | 1 | New York | 1 |
| Johns Hopkins | 13 | Minnesota | 4 | Tufts | 1 | New Mexico | 1 |
| | | | | | | Utah | 1 |

*RESEARCH ORGANIZATIONS*

| U. S. Government: | | Endowed and Industrial Research Groups: | | | |
|---|---|---|---|---|---|
| Geological Survey | 8 | Carnegie Inst. Wash. | 27 | Ortho Research Found. | 1 |
| Smithsonian | 3 | Rockefeller Inst | 19 | Research Corp. | 1 |
| Bur. Standards | 3 | Inst. Advanced Study | 7 | Wistar Inst. | 1 |
| Nat. Inst. Health | 3 | Am. Tel. and Tel. | 6 | Bethlehem Steel Co. | 1 |
| Atomic Energy Comm. | 2 | General Electric | 6 | Carbide & Carbon Chem. | 1 |
| Adv. Com. Aeronautics | 2 | Am. Mus. Nat. Hist. | 2 | E. I. du Pont de Nemours | 1 |
| Bur. Entomology | 1 | N. Y. Botan. Garden | 2 | General Motors | 1 |
| Bur. Plant Indust. | 1 | Conn. Agr. Station | 2 | International Nickel | 1 |
| Research and Development Board. | 1 | Gulf Res. & Dev. | 2 | Lederle Laboratories | 1 |
| | | A.A.A.S. | 1 | Merck and Co. | 1 |
| Weather Bureau | 1 | R. B. Jackson Mem. Lab. | 1 | Monsanto Chemical Co. | 1 |
| | | | | Pan Am. Petroleum | 1 |
| | | Loomis Lab. | 1 | Radio Corp. America | 1 |
| | | Lowell Observatory | 1 | Univ. Oil Products | 1 |
| | | N. J. Agr. Exp. Sta. | 1 | Westinghouse Electric | 1 |
| | | N. Y. State Museum | 1 | J. G. White Corp. | 1 |

Table 4 is supplemental to table 3 and shows the range of ages at election (column I) and as of July 1, 1948 (column II), in each section. This table is of special interest because it indicates that workers in certain fields of science, such as mathematics, physics, chemistry and astronomy (the physical sciences) develop and establish records of accomplishment earlier than is the case with workers in the sciences which are in part descriptive in character and require field work (geology, botany, engineering) or long experience in the behavior of living matter (biological sciences). In the case of psychology, a relatively new science, many workers are applying the methods of the physical and biological sciences to the study of the behavior of the human being and some of its younger workers have been elected to Academy membership. The last line of the table presents in percentage

TABLE 3

AGES OF MEMBERS AT ELECTION AND ON JULY 1, 1948

| | | | AVERAGE AGES OF SECTION GROUPS | | | |
|---|---|---|---|---|---|---|
| YEARS | NO. AT ELECTION | NO. ON JULY 1, 1948 | NAME OF SECTION | NO. IN SECTION | AGE AT ELECTION | AGE ON JULY 1, 1948 |
| 30–34 | 9 | 0 | 1. Mathematics | 26 | 43.1 | 57.9 |
| 35–39 | 40 | 7 | 2. Astronomy | 26 | 47.8 | 67.9 |
| 40–44 | 91 | 34 | 3. Physics | 60 | 44.8 | 57.0 |
| 45–49 | 98 | 47 | 4. Engineering | 35 | 52.5 | 62.4 |
| 50–54 | 96 | 43 | 5. Chemistry | 58 | 47.3 | 57.8 |
| 55–59 | 52 | 68 | 6. Geology | 38 | 54.0 | 67.2 |
| 60–64 | 30 | 69 | 7. Botany | 35 | 50.8 | 62.3 |
| 65–69 | 17 | 56 | 8. Zoology | 38 | 51.3 | 62.7 |
| 70–74 | 2 | 51 | 9. Physiology | 45 | 49.7 | 59.1 |
| 75–79 | 0 | 31 | 10. Pathology | 37 | 53.3 | 63.2 |
| 80–84 | 0 | 20 | 11. Anthropology | 9 | 52.6 | 63.6 |
| 85–89 | 0 | 5 | 12. Psychology | 23 | 47.9 | 62.0 |
| 90–94 | 0 | 4 | | | | |

| | | | |
|---|---|---|---|
| Average age July 1, 1948 | 61.5 | Median age July 1, 1948 | 61.1 |
| Average age at election | 49.2 | Median age at election | 48.9 |
| Average age at election (1935) | 48.9 | | |

TABLE 4

RANGES OF AGES IN EACH SECTION AT ELECTION (COL. I) AND ON JULY 1, 1948 (COL. II)

| | | AGE RANGE IN 10-YEAR PERIODS | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 30–40 | | 40–50 | | 50–60 | | 60–70 | | 70–80 | | 80–90 | | 90–100 | |
| SECTION | NO. IN SECTION | (I) | (II) | (I) | (II) | (I) | (II) | (I) | (II) | (I) | (II) | (I) | (II) | (I) | (II) |
| 1. Mathematics | 26 | 9 | 1 | 14 | 7 | 3 | 6 | 0 | 8 | 0 | 3 | 0 | 1 | 0 | 0 |
| 2. Astronomy | 26 | 4 | 0 | 12 | 2 | 7 | 3 | 3 | 7 | 0 | 12 | 0 | 2 | 0 | 0 |
| 3. Physics | 60 | 14 | 1 | 34 | 25 | 9 | 10 | 3 | 13 | 1 | 8 | 0 | 3 | 0 | 0 |
| 4. Engineering | 35 | 2 | 1 | 9 | 3 | 19 | 11 | 4 | 12 | 1 | 6 | 0 | 2 | 0 | 0 |
| 5. Chemistry, | 58 | 10 | 3 | 28 | 12 | 16 | 19 | 3 | 16 | 0 | 4 | 0 | 4 | 0 | 0 |
| 6. Geology, | 38 | 1 | 0 | 10 | 4 | 16 | 8 | 11 | 11 | 0 | 9 | 0 | 4 | 0 | 2 |
| 7. Botany | 35 | 0 | 0 | 20 | 6 | 9 | 10 | 6 | 8 | 0 | 9 | 0 | 1 | 0 | 1 |
| 8. Zool. & Anat. | 38 | 1 | 0 | 19 | 4 | 13 | 16 | 5 | 8 | 0 | 7 | 0 | 3 | 0 | 0 |
| 9. Physiol. & Bioch. | 45 | 3 | 1 | 21 | 8 | 17 | 13 | 4 | 17 | 0 | 6 | 0 | 0 | 0 | 0 |
| 10. Pathol. & Bact. | 37 | 1 | 0 | 9 | 5 | 22 | 8 | 5 | 15 | 0 | 7 | 0 | 2 | 0 | 0 |
| 11. Anthropology | 9 | 0 | 0 | 3 | 1 | 5 | 3 | 1 | 2 | 0 | 3 | 0 | 0 | 0 | 0 |
| 12. Psychology | 23 | 3 | 0 | 9 | 5 | 10 | 4 | 1 | 7 | 0 | 6 | 0 | 1 | 0 | 0 |
| Total | 430 | 48 | 7 | 188 | 82 | 146 | 111 | 46 | 124 | 2 | 80 | 0 | 23 | 0 | 3 |
| Per cent | | 11 | 2 | 44 | 19 | 34 | 26 | 11 | 29 | 1 | 18 | 0 | 5 | 0 | 1 |

form the distribution of ages at election and on July 1, 1948, of the section members in the ten-year ranges.

In table 5 the numbers of members, as of July 1, 1948, who were elected in a given year, are listed. In the table the effect of raising in 1942 the limit of the number of members who may be elected at any annual meeting is

TABLE 5

NUMBER OF PRESENT MEMBERSHIP WHO WERE ELECTED IN A GIVEN YEAR

| YEAR | NO. ELECTED | YEAR | NO. ELECTED | YEAR | NO. ELECTED | YEAR | NO. ELECTED |
|---|---|---|---|---|---|---|---|
| 1948 | 30 | 1938 | 15 | 1928 | 8 | 1918 | 7 |
| 1947 | 29 | 1937 | 13 | 1927 | 10 | 1917 | 7 |
| 1946 | 28 | 1936 | 11 | 1926 | 0 | 1916 | 4 |
| 1945 | 29 | 1935 | 10 | 1925 | 7 | 1915 | 3 |
| 1944 | 24 | 1934 | 11 | 1924 | 7 | 1914 | 1 |
| 1943 | 25 | 1933 | 12 | 1923 | 7 | 1913 | 4 |
| 1942 | 15 | 1932 | 11 | 1922 | 7 | 1912 | 1 |
| 1941 | 15 | 1931 | 9 | 1921 | 6 | 1911 | 1 |
| 1940 | 13 | 1930 | 11 | 1920 | 8 | 1910 | 3 |
| 1939 | 14 | 1929 | 11 | 1919 | 5 | 1909 | 1 |
| | | | | | | 1908 | 2 |
| Total | 222 | | 114 | | 65 | | 34 |

TABLE 6

NUMBER OF ACADEMY MEMBERS, AS OF JULY 1, 1948, IN EACH SECTION AND AGES OF ELECTION WITHIN EACH 10-YEAR RANGE. IN EACH GROUP COLUMN I LISTS NUMBER OF MEMBERS ELECTED BEFORE 1943; COLUMN II, THOSE ELECTED THEREAFTER

| | AGE RANGE | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 30–40 | | 40–50 | | 50–60 | | 60–70 | | 70–80 | |
| SECTION | (I) | (II) | (I) | (II) | (I) | (II) | (I) | (II) | (I) | (II) |
| 1. Mathematics | 6 | 3 | 10 | 4 | 3 | 0 | 0 | 0 | 0 | 0 |
| 2. Astronomy | 4 | 0 | 11 | 1 | 7 | 0 | 2 | 1 | 0 | 0 |
| 3. Physics | 11 | 3 | 16 | 18 | 5 | 4 | 1 | 2 | 0 | 0 |
| 4. Engineering | 1 | 1 | 5 | 4 | 10 | 9 | 3 | 1 | 1 | 0 |
| 5. Chemistry | 5 | 5 | 18 | 10 | 5 | 11 | 0 | 3 | 1 | 0 |
| 6. Geology | 1 | 0 | 6 | 4 | 9 | 7 | 7 | 4 | 0 | 0 |
| 7. Botany | 0 | 0 | 10 | 10 | 7 | 2 | 3 | 3 | 0 | 0 |
| 8 Zool. & Anat. | 1 | 0 | 12 | 7 | 6 | 7 | 4 | 1 | 0 | 0 |
| 9. Physiol. & Bioch. | 1 | 2 | 13 | 8 | 6 | 11 | 0 | 4 | 0 | 0 |
| 10. Pathol. & Bact. | 1 | 0 | 6 | 3 | 13 | 9 | 0 | 5 | 0 | 0 |
| 11. Anthropology | 0 | 0 | 3 | 0 | 3 | 2 | 0 | 1 | 0 | 0 |
| 12. Psychology | 2 | 1 | 6 | 3 | 8 | 2 | 0 | 1 | 0 | 0 |
| Total | 33 | 15 | 116 | 72 | 82 | 64 | 20 | 26 | 2 | 0 |
| Per cent | 13 | 8 | 46 | 41 | 32 | 36 | 8 | 15 | 1 | 0 |

clearly shown. In this list the members emeriti are included. The list shows that the period of service of 222 members is from 0 to 10 years; of 114 members, 10 to 20 years; of 65 members, 20 to 30 years; of 32 members 30 to 40 years; and of 2 members, in excess of 40 years.

In table 6 the numbers (as of July 1, 1948) of members elected prior to

1943 are listed in the first (I) of two columns in each section and for each ten-year range of age at election; in the second column (II) similar data for numbers of section members elected to membership between 1943 and 1948 are presented. At the foot of each column its sum is given; the total number of section members elected before 1943 is 253; that for members elected thereafter is 177. In the last line the ratios are expressed in percentages; thus in column I of the age range 30–40 the number 33 is 13 per cent of 253; the number 15 of column II is 8 per cent of 177. Comparison of the percentages shows perceptible shifts in the age groups, 30–40 and 60–70, as a result of the change from 15 to 30 in the limit of the candidates who may be elected each year. The elections of 1943 to 1948 are notable for their low percentage in the lower and upper age groups. Too great conservatism with respect to the lower age group may be just as bad as too little and may reduce the effectiveness of the Academy, especially in its relation to Government problems. Younger men have the energy, enthusiasm and capacity to undertake the solution of a given problem and to do this in addition to their ordinary tasks. Many men on reaching 70 are less active than in earlier years and cannot undertake large tasks. In the Academy the number of members past 70 is 106, in other words nearly one quarter of the entire membership has attained the usual retiring age. The number of Academy members whose ages exceed 60 is 230, a number greater than one-half the Academy membership.

TABLE 7

AGES AT ELECTION FOR THE PERIODS: (I) 1863–1924; (II) 1925–1942; (III) 1943–1948; (IV) 1863–1948

| ELECTED: AGE | 1863–1924 (I) | 1925–1942 (II) | 1943–1948 (III) | 1863–1948 (IV) | AGE | 1863–1924 (I) | 1925–1942 (II) | 1943–1948 (III) | 1863–1948 (IV) |
|---|---|---|---|---|---|---|---|---|---|
| 26–27 | 1 | 0 | 0 | 1 | 56–57 | 17 | 8 | 9 | 34 |
| 28–29 | 1 | 0 | 0 | 1 | 58–59 | 17 | 10 | 9 | 36 |
| 30–31 | 4 | 1 | 0 | 5 | 60–61 | 7 | 9 | 8 | 24 |
| 32–33 | 7 | 4 | 0 | 11 | 62–63 | 16 | 12 | 4 | 32 |
| 34–35 | 15 | 4 | 2 | 21 | 64–65 | 6 | 11 | 4 | 21 |
| 36–37 | 29 | 5 | 5 | 39 | 66–67 | 6 | 5 | 4 | 15 |
| 38–39 | 25 | 11 | 6 | 42 | 68–69 | 7 | 2 | 5 | 14 |
| 40–41 | 29 | 17 | 13 | 59 | 70–71 | 3 | 3 | 0 | 6 |
| 42–43 | 28 | 9 | 17 | 54 | 72–73 | 3 | 1 | 0 | 4 |
| 44–45 | 39 | 17 | 16 | 72 | 74–75 | 2 | 0 | 0 | 2 |
| 46–47 | 32 | 18 | 11 | 61 | 76–77 | 0 | 0 | 0 | 0 |
| 48–49 | 35 | 22 | 13 | 70 | 78–79 | 0 | 0 | 0 | 0 |
| 50–51 | 38 | 38 | 15 | 91 | 80–81 | 0 | 1 | 0 | 1 |
| 52–53 | 27 | 14 | 16 | 57 | 82–83 | 1 | 0 | 0 | 1 |
| 54–55 | 26 | 16 | 12 | 54 | | | | | |
| | | | | | Total | 421 | 238 | 169 | 828 |

Of the 407 members elected during the period 1925–1948 the number living on July 1, 1948, was 359; of these, the entire number (38) of members whose age at election was less than 40 is still living. The average age at election of the 359 members is 50.0. The average election age of the 47 members who have died is 58.7 years; the list includes 3 members elected at the ages: 80.2, 72.5 and 70.1, respectively; 20 between 60 and 70; 17 between 50 and 60; and 6 between 40 and 50.

*Vital Statistics of Members Elected during the Period 1925–1948.*—Comparison can now be made of the ages of members elected during the years 1925 to 1948 with those of the 421 members elected during the period 1863–1924, including charter members, as reported upon by Pearl in 1925. The necessary data are presented in tables 7 and 8. In table 7 the numbers of members elected in each age group are listed for the four periods: 1863–1924 (column I as given by Pearl); 1925–1942 (column II); 1943–1948 (column III); and 1863–1948 (column IV). In each column all members (living, resigned and deceased) who were elected during the given period are included. If a graph be prepared, one curve for each column I, II, III, inspection of the curves shows that the tendency to elect older men has increased in the course of the years. The increase may be connected with that of the life span during the past several generations; to the membership, which now elects the new members, a man of 40 may appear to be of the same degree of youngness as did a man of 30 or 35 two or three generations ago. Another factor in favor of the rise may be the present tendency toward regimentation and coöperative effort in solution of problems under group attack. At present this is felt in all walks of life and may tend to retard recognition of contributions by young investigators and to have a lag effect on the ages at election of the Academy membership.

TABLE 8

VITAL STATISTICS ON AGES AT ELECTION

| PERIOD | MEAN AGE | MEDIAN AGE | PER CENT UNDER 40 | NO. ELECTED | AVERAGE EACH YEAR |
|---|---|---|---|---|---|
| Charter 1863 | 51.7 | 51.3 | 16.7 | 48 | . . . |
| 1863–1883 | 44.5 | 41.3 | 45.5 | 95 | 5 |
| 1884–1904 | 46.5 | 45.1 | 20.0 | 65 | 3 |
| 1905–1924 | 50.5 | 49.5 | 8.4 | 213 | 11 |
| 1925–1942 | 51.0 | 50.6 | 10.5 | 238 | 13 |
| 1943–1948 | 50.6 | 50.2 | 7.7 | 169 | 28 |

In 1945 Edwin B. Wilson compared the results obtained by Pearl on the ages at election for certain periods covering the years 1863 to 1924 with those for the years 1943 to 1945. He concluded with Pearl "that the tendencies toward a higher mean age, a median nearer the mean, . . . have persisted and been intensified." On the basis of the data considered by Wilson and presented in his table, this conclusion was correct and justified. However, as shown in table 8, the addition of data for the periods 1925–1942 and 1943–1948 indicates that for the periods 1905–1924, 1925–1942, 1943–1948, the average ages at election, the percentages of younger men elected at ages under 40 and of older men elected at 62 and at a later age have remained nearly constant. Thus, for the entire period 1905–1948, the average age at election was 50.8; for the years 1904–1924 it was 50.5; for 1925–1942, 51.0; and for 1943–1948, 50.6. The ratios of men under 40 elected during the same periods were, respectively: 56 to 620 or 9.0 per cent; 18 to 213 or 8.5 per cent; 25 to 238 or 10.5 per cent; and 11 to 169 or 7.7 per cent. Similarly for men 62 and older at election the numbers for the same periods were, respectively: 75 to 620 or 12.1 per cent; 23 to 213 or 10.8 per cent; 35 to 238 or 14.7 per cent; 17 to 169 or 10.1 per cent.

These results indicate that for the past 44 years the Academy membership with respect to ages at election has changed but little. For shorter periods, such as a single year or only a few years the ratios may vary appreciably. Thus for the 6-year period 1943–1948 a slight decrease in the average age and a perceptible decrease in the percentage of the men elected under 40 are noticeable. In spite of the small fluctuations the figures indicate that the average age at election and the percentages of men elected at ages under 40 and 62 or older have remained reasonably constant; in other words, the Academy membership has continued to be stable and in balance.

---

# *THE MIOCENE OCCURRENCE OF SEQUOIA AND RELATED CONIFERS IN THE JOHN DAY BASIN*

BY RALPH W. CHANEY

DEPARTMENT OF PALEONTOLOGY, UNIVERSITY OF CALIFORNIA, AND RESEARCH ASSOCIATE, CARNEGIE INSTITUTION OF WASHINGTON

Read before the Academy, November 16, 1948

Four genera of the family Taxodiaceae—Sequoia, Taxodium, Metasequoia and Glyptostrobus—are among the most numerous and widely distributed conifers in the Tertiary record of Western North America. All of them have been found in rocks of Middle Miocene age in the John Day Basin and adjacent areas of eastern Oregon, although none of them live there today. Taxodium and Metasequoia occur together in the Mascall flora, and the other two are members of related floras located within sixty miles.[1] By contrast the distribution of existing trees of these genera shows no overlapping ranges; Sequoia and Taxodium are confined to North America while Metasequoia and Glyptostrobus now live only in Asia. *Sequoia sempervirens* is found along the Pacific Coast from central California to Oregon;[2] *Taxodium distichum* occupies the Atlantic coastal plain from Delaware south to Florida, and around the Gulf to the Mexican border, with a northward extension up to the Mississippi valley to southern Illinois and Indiana; a second species, *T. mucronatum*, extends from northern Mexico into Guatemala. *Metasequoia glyptostroboides* is known to occur in a limited area in central China, near the Szechuan-Hupeh border; *Glyptostrobus pensilis* has a restricted distribution in the lowlands of southern China.[3] It is apparent that during the approximately twenty million years since the Middle Miocene, events have transpired, changes have occurred, which have eliminated all of these conifers from the John Day Basin, and have greatly restricted their distribution elsewhere.

All four of the genera under consideration have been recorded from the Cretaceous[4] and Eocene[5] deposits of Alaska, or from Greenland, Iceland, Spitzbergen, arctic Siberia and the islands north of Canada.[6] At these northern localities they are associated with hardwood-deciduous genera such as Acer, Alnus, Betula, Carpinus, Castanea, Cercidiphyllum, Fagus, Liquidambar, Quercus and Ulmus. This mixed conifer-hardwood assemblage, which appears to have had its origin at high northern latitudes, is known as the Arcto-Tertiary Flora. Up to the close of the Eocene epoch, approximately forty million years ago, mild climates made possible the existence of these trees far beyond their present northern limits. At about that time there is evidence of a progressive change toward lower winter temperatures which continued down to the Pleistocene, resulting in restriction of the ranges of most of these trees at the north, and their extension into lower latitudes. By the close of the Oligocene epoch, some ten million years later, the Arcto-Tertiary Flora was widely established at middle latitudes, where the fossil record of its presence is known from Oregon, Manchuria and elsewhere on both sides of the Pacific. *M. heerii* was the dominant conifer of this forest in eastern Oregon, and judging from the habits of its living equivalent, *M. glyptostroboides*, in central China, it occupied moist valleys and ravines at moderate elevations, under climatic conditions characterized by humid summers and mild winters.[7] *T. dubium* was an uncommon associate, as judged by its fossil representation; this tree, like its living equivalent, may be considered to have lived in valley swamp habitats. The other two related genera are not known to have lived nearby, though *S. affinis* is of common occurrence in Oligocene deposits near the coast of Oregon and Washington, and both it and *G. oregonensis* have been recored from localities farther inland.

By the middle of the Miocene epoch which followed, changes were taking place which had a profound effect upon the topography and climate of the John Day Basin and upon its vegetation. The Cascade Range was being built up by lava flows and accumulations of pyroclastic materials,[8] and streams flowing westward across its axis were being dammed or diverted. Local eruptions in eastern Oregon were further altering the drainage, producing many small lakes and swamps along the stream courses.

Deposits of volcanic ash, which make up so conspicuous a part of the Mascall formation, were accumulating in these basins, and were enclosing one of the most complete known records of the vegetation of the past. Along the poorly drained valleys of the John Day area, the swamp cypress, *T. dubium*, took the place of *M. heerii* as the dominant conifer, together with a group of hardwoods whose living equivalents are now associated with the swamp cypress in the eastern United States, including such genera as Acer, Betula, Carya, Nyssa, Platanus and Quercus. Metasequoia persisted as a member of the forest on higher ground around the borders, as judged by the occurrence of its leafy shoots and cones in relatively small numbers in the fossil record.

From two localities in the Blue Mountains, some 40 miles to the east, a flora has been collected which shows many species in common with the Mascall flora, but which includes a group of upland trees not therein represented.[9] From this Blue Mountains flora both *T. dubium* and *M. heerii* are missing. The common occurrence of Sequoia, a genus now confined to Coastal valleys and slopes up to elevations of 3000 feet, in the interior of Oregon is not fully explained; it has been recorded also from somewhat older floras still farther inland, the Missoula flora of Montana, the Intermediate flora of Yellowstone Park, the Florissant flora of Colorado, among groups of plants which appear to have occupied upland environments. Apparently Sequoia was finding it difficult to survive under the new climatic regime, and prior to its extinction in the interior was making a last stand at higher elevations where there was adequate precipitation. The ranges of several existing species of conifers suggest a similar contemporary adjustment to dry interior conditions; *Abies grandis, Pinus contorta, Pseudotsuga taxifolia, Taxus brevifolia, Thuja plicata* and *Tsuga heterophylla* show discontinuous distribution from the Pacific Coast into the mountains as far east as Montana; *Chamaecyparis lawsoniana, C. nootkatensis* and *Picea sitchensis* extend eastward for lesser distances. The environment of the Blue Mountains flora, as shown by Sequoia and other conifers, was clearly of a more upland character than that of the Mascall flora, with its dominant Taxodium and associated swamp types of hardwoods.

The occurrence of *G. oregonensis* as the dominant conifer of the Stinking Water flora, in the Stinking Water Basin, 34 miles east of Burns on the south side of the Blue Mountains, is somewhat difficult to interpret, especially since there is little available information regarding the natural environment of trees of this genus in southern China. The water pine, *G. pensilis*, is reported to grow on open stream courses and sandy delta margins near sea level. It appears improbable that a similar coastal environment was available in the Stinking Water Basin during the Miocene epoch, at which time the nearest recorded strand was nearly 300 miles to the west. The general aspect of the Stinking Water flora, and of the Beulah and Succor Creek floras immediately to the east, both of which also include Glyptostrobus, suggests more open country than that in the John Day Basin, with probably much less rainfall. On this terrain, largely occupied by an oak woodland, the Miocene water pines appear to have lived along the borders of large streams, in whose deposits their leafy shoots and cones have accumulated and have been preserved in great abundance.

Although the proximity of the four taxodiaceous genera during the Miocene is not wholly explained by these suggestions, it seems apparent that each of them occupied a distinctive ecologic niche, and that the topography in this part of eastern Oregon was of a varied character, with swampy valleys, cool moist slopes and broad, open savannas traversed by large streams. The deciduous habit of all except Sequoia, together with the predominantly deciduous habit of the hardwood associates, may be interpreted as indicating a summer-wet climate, and a cold though probably not severe winter. In preceding Eocene time, before the Cascade Range had reached sufficient height to constitute a climatic barrier, the subtropical type of vegetation which characterized the coastal margin as known from the Goshen and Comstock floras was also represented in the Clarno flora of the John Day Basin. During the Oligocene epoch which followed, there was already apparent a differentiation between the coastal vegetation, with Sequoia, and that of the interior, with Metasequoia. The Miocene epoch shows still further changes in the interior, with a trend toward the continental type of climate which now characterizes this region. The rich mixed forest of conifers and hardwoods gradually disappeared as the Cascade barrier to the west continued to rise, reducing precipitation to its present level of some eleven inches a year in the John Day Basin, and bringing lower winter temperatures. By Pliocene time only the hardier members of the Arcto-Tertiary Flora were living east of the Cascades; today a few of them—birch, cherry, fir, maple, pine and poplar—remain at higher levels or along streams where there is a permanent supply of water. Sequoia, Taxodium, Metasequoia and Glyptostrobus have long since disappeared.

With the exception of Sequoia, all of these genera now live in regions with summer rainfall; heavy summer fogs supplement the moisture requirements of Sequoia, which lives along the Pacific Coast in a region of winter rainfall. Mild temperatures characterize the habitats of all four genera; although Taxodium ranges north up the Mississippi Basin to southern Illinois, where winter temperatures fall well below freezing for several months, the swamp cypress attains its largest size and greatest abundance in the southerly portion of its range. It seems probable that throughout its history the Arcto-Tertiary Flora, made up largely of deciduous conifers and hardwoods, has been largely confined to areas characterized by summer rainfall, and that it has had its best development where seasonal changes in temperature have not been extreme. Elements of this Flora are still widely distributed at middle latitudes in the northern hemisphere, but only locally, where their delicately balanced climatic and topographic requirements are fully met, have Sequoia, Taxodium, Metasequoia and Glyptostrobus survived. The continued existence of these genera is dependent upon the extent of further environmental changes, as well as upon man's wisdom in conserving them.

---

[1] A study of the Miocene floras of the Columbia Plateau, by D. I. Axelrod and the writer is nearing completion.

[2] A related genus, Sequoiadendron, occurs in the Sierra Nevada of California.

[3] Other genera of the Taxodiaceae. Mention of other genera of the Taxodiaceae is here omitted, since none of them is known to occur in abundance in the fossil record.

[4] Hollick, A., *U. S. Geol. Surv.*, Prof. Paper 159, 1930.

[5] Hollick, A., *U. S. Geol. Surv.*, Prof. Paper 182, 1936.

[6] Heer, O., *Flora Foss. Arct.*, 7 vols. (1868–1883).

[7] Chaney, R. W., these PROCEEDINGS, **34**, 503–515(1948).

[8] Hodge, E. T., *Pan-Am. Geol.*, **49**, 341–356 (1928).

[9] Oliver, E., *Carnegie Inst. Washington*, Pub. 455, 1–27 (1934).

---

# *A NOTE ON CENTROMERE ORGANIZATION*

BY M. M. RHOADES AND WARWICK E. KERR

DEPARTMENT OF BOTANY, UNIVERSITY OF ILLINOIS, URBANA, AND THE SECTION OF GENETICS, UNIVERSITY OF SAO PAULO, PIRACICABA, BRAZIL

Communicated January 11, 1949

The centromere or kinetochore is that portion of the chromosome concerned with congression to and orientation on the spindle at metaphase as well as with the more dramatic poleward movement at anaphase. In the great majority of plants and animals the centromere is a sharply delimited region of the chromosome differing in its morphology, chemical constitution and function. During anaphase the centric region leads the poleward advance since the half-spindle components which bring about the movement are elaborated solely by the localized centromere, and the chromosome arms play no active rôle, being passively pulled along the spindle to the poles. Acentric fragments form no spindle fibres and their behavior on the spindle figure is erratic.

There exist, however, chromosomes with half-spindle fibres originating not solely from a localized region of the chromosome but from the whole length of the chromosome. In these chromosomes the entire poleward face of the chromosome is involved in fibre formation and a solid sheet of spindle fibres is produced. Schrader[1] found that this diffuse type of centromere characterized the chromosomes of the hemipterous and homopterous insects. Chromosomes with diffuse centromeres are oriented on the spindle so that the body of the chromosome lies in a plane perpendicular to the longitudinal axis of the spindle, and the separating chromosomes maintain a position parallel to the spindle plate, although in some species the ends of the chromosomes are bent towards the pole (this is clearly an adaptation allowing the body of the chromosome to converge to the pole of the spindle). As Hughes-Schrader and Ris[2] and Ris[3] have shown, the behavior of fragments derived from chromosomes with diffuse centromeres is different from that of those derived from chromosomes with localized centromeres. They observed that these fragments were oriented on the metaphase spindle in a normal manner; each fragment formed spindle fibres along its length and had the broad-side orientation on the spindle characteristic of intact chromosomes. These fragments behaved normally in a number of mitoses. It appears in this type of centromere that the ability to form spindle fibres is shared by all regions of the chromosome—i.e., the centromere is spread or diffused continuously along the chromosome.

In addition to chromosomes with single, localized centromeres and those with diffuse centromeres is the polycentric type found in the germ line of *Ascaris megalocephala* where the multiple or compound chromosomes consist of an end-to-end union of smaller chromosomes each possessing a localized centromere. The anaphase movement of the compound chromosomes is the resultant of the action of the individual centromeres. White[4] found that experimentally induced fragments of these compound chromosomes were normally oriented on the spindle. The polycentric and diffuse types have a superficial similarity but they differ in a fundamental manner since the centromere substance is believed to be spread continuously along the chromosome in diffuse centromeres while the polycentric chromosomes possess many small individual centromeres. Presumably every fragment, no matter how minute, from a chromosome with a diffuse centromere would possess centric material and be able to form fibres while polycentric chromosomes could not be subdivided indefinitely into normally behaving fragments.

Among plants chromosomes with other than localized centromeres have been rarely observed, but Malheiros, de Castro and Câmara[5] found nonlocalized centromeres in the rush *Luzula purpurea.* It is not clear whether these chromosomes are polycentric or have diffuse centromeres, but a polycentric condition appears much more likely. Further, there are, both in maize and rye, strains in which regions of the chromosomes other than the true centromere form spindle fibres (Rhoades and Vilkomerson,[6] Prakken and Müntzing,[7] and Östergren and Prakken[8]). This unusual behavior is restricted to the two meiotic divisions, and in maize we have been able to demonstrate that the formation of these subsidiary spindle fibres is a product of the activity of the true centric region (Rhoades[9]).

There is also the dicentric condition reported by Piza[10] for the chromosomes of the Brazilian scorpion *Tityus bahiensis* ($2n = 6$) where, according to Piza, each of these large chromosomes has a terminal centromere at both ends. He came to this conclusion because the two ends of the chromosomes lead the way to the pole even though spindle fibres are present along the entire poleward face of the chromosome. These interjacent fibres he believed arose from some influence emanating from the two end centromeres and moving along the body of the chromosome. Piza also holds that the diffuse centromeres of the insect orders studied by Schrader are of the same dicentric type as in Tityus since in some of these forms the ends also pass to the poles in advance of the rest of the chromosome. However the end centromeres postulated by Piza have never been observed and there are reasons for believing that the Tityus chromosomes are polycentric. Whether or not these chromosomes of the scorpion are dicentric or polycentric can be determined by a simple experimental procedure—i.e., the fracturing of the chromosome by short-wave irradiation. We, therefore, treated nine Tityus males with x-rays, their testes were removed at various times following treatment, fixed in Dusbock-Brasil fluid, sectioned and stained with hemotoxylin.* The discrimination between the dicentric and polycentric condition can be obtained from the behavior of fragments with recently broken ends. As Piza has correctly pointed out, the ends of the chromosom espass to the pole precociously (this is first evident at late metaphase); this he attributes to the terminal centromeres. However, the turning of the ends towards the poles may be nothing more than an adaptation permitting the chromosomes (huge compared to cell size) to converge to the spindle poles and they may be polycentric. If the intact chromosomes possess two terminal centromeres then fragments would at most have only one normal end while the other end must be a freshly broken one; it would be expected then that the normal end would continue to show a precocious poleward movement while the freshly broken end would have an anaphase migration no different from the rest of the chromosome. On the

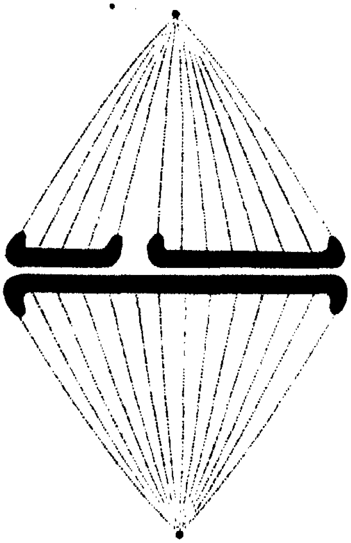


FIGURE 1

A bivalent at late metaphase in a primary spermatocyte of *Tityus bahiensis*. The upper homolog has been fractured as the consequence of x-ray irradiation. The two freshly broken ends show the same precocious poleward movement as do the normal ends.

*other hand, if these scorpion chromosomes are polycentric the old and new ends might both show the same early poleward advance and no difference* in behavior on the spindle would be manifest. This is precisely what was observed in several score of primary and secondary spermatocytes with fragmented chromosomes. One such case is illustrated in figure 1. There is, in our opinion, no convincing evidence that Tityus has dicentric chromosomes. Rather the behavior of the chromosome fragments reported above together with the fact that close relatives of Tityus have numerous small chromosomes with localized centromeres and that the postulated terminal centromeres have never been observed make it appear highly probable that these chromosomes are polycentric.

Centromeres have been classed as either localized or non-localized. The diffuse centromere found in the hemiptera and homoptera, where it is apparently spread continuously along the chromosome, is the best example of the non-localized type. The proper classification of polycentric chromosomes is somewhat uncertain on this basis. Their appearance and behavior on the spindle is hardly distinguishable from that of the hemipteran chromosomes with diffuse centromeres yet these polycentric chromosomes are constituted by an end-to-end union of numerous small chromosomes each with its own individualized centromere. Chromosomes with localized centromeres such as are found in the great majority of organisms are monocentric. The evidence for dicentric chromosomes in Tityus is unconvincing; the experimental evidence suggests they are polycentric. It is to be questioned if dicentric chromosomes ever are found in any normal chromosome complement. Of some interest in connection with the problem of the evolution of centromere form and function is the situation obtaining in those strains of maize and rye where normally monocentric chromosomes have subsidiary fibres at regions devoid of centromeres. However, in maize it has been ascertained that the formation of these extra fibres is controlled, in part at least, by the localized centromere.

---

[*] We wish to express our gratitude to Professor André Dreyfus of the University of São Paulo, São Paulo, for making the arrangements necessary for the x-ray treatments and to Professor F. G. Brieger for providing facilities.

[1] Schrader, F., *Cytologia*, **6**, 422–431 (1935); *Mitosis*, Columbia University Press, 1944.

[2] Hughes-Schrader, S., and Ris, Hans, *J. Exp. Zool.*, **87**, 429–456 (1941).

[3] Ris, Hans, *Ibid.*, **90**, 267–330 (1942).

[4] White, M. J. D., *Nature*, **137**, 783 (1936).

[5] Malheiros, N., de Castro, D., and Câmara, A., *Agronomia Lusitana*, **9**, 51–74 (1947).

[6] Rhoades, M. M., and Vilkomerson, H., these PROCEEDINGS, **28**, 433–436 (1942).

[7] Prakken, R., and Müntzing, A., *Hereditas*, **28**, 441–481 (1942).

[8] Östergren, G., and Prakken, R., *Ibid.*, **32**, 473–494 (1946).

[9] Rhoades, M. M. (unpublished).

[10] Piza, S. de T., *Sci. Genet.*, **1**, 255–261 (1939); *Ann. Esc. Sup. Agric. "Luiz de Queiroz,"* **3**, 27–54 (1946); *Ibid.*, **3**, 55–67 (1946).

# PRIMARY IDEALS IN GROUP ALGEBRAS*

By Irving Kaplansky

Institute for Advanced Study

Communicated by M. H. Stone, January 29, 1949

Let $G$ be a locally compact abelian group and $A$ its $L_1$-algebra, that is, the algebra of all complex functions summable with respect to Haar measure, with convolution as multiplication. Let $I$ be a closed ideal in $A$. A question stemming from Wiener's Tauberian investigations is the following: is $I$ the intersection of the regular maximal ideals containing it? L. Schwartz[1] has recently published a counter-example. This negative result increases the interest in special cases where the answer is affirmative. Ditkin,[2] Beurling[3] and Segal[4] have obtained an affirmative answer in the following case: $I$ is a primary ideal (that is, it is contained in exactly one regular maximal ideal), and $G$ is the group of real numbers or the group of integers. The purpose of this note is to extend this result to any locally compact abelian group. This will be accomplished by the methods of Segal, together with the structure theorems for groups. The essential idea is that the discrete case can be treated "elementwise," and thus reduced to the case of a cyclic group.

THEOREM 1. *In the $L_1$-algebra of a locally compact abelian group, any closed primary ideal is maximal.*

In the proof we shall use the following lemma, which is a slight extension of Lemma 2.7.1 in S.

LEMMA 1. *Let $R$ be the direct sum of a finite number of groups, each of which is the group of real numbers or of integers. Then in $L_1(R)$ there exists a bounded sequence $f_n$ such that (1) the Fourier transform of each $f_n$ is 1 in a neighborhood of 0 (the neighborhood depending on $n$), (2) $f_ng - f_n\int_R g \to 0$ as $n \to \infty$, for any $g$ in $L_1(R)$.*

Any locally compact abelian group $G$ can be written as the direct sum of a vector group $V$ and a group $H$ having a compact open subgroup $C$. We use the additive notation, and the symbol $O$ will denote the identity element of any group. Also we shall write $(f, g)$ for the element of $L_1(G)$ which is the product of $f$ in $L_1(V)$ and $g$ in $L_1(H)$. Let $I$ be a closed primary ideal in $L_1(G)$ and $M$ the unique regular maximal ideal containing it. Our task is to prove $I = M$. We may suppose without loss of generality that the character associated with $M$ is $O$. The following lemma constitutes the essential step in the proof.

LEMMA 2. *Let $e$ be the characteristic function of $C$, and let $p$, $q$ be arbitrary elements in $L_1(V)$, $L_1(H)$; write $c = \int_H q$. Then $(p, q - ce)$ is in $I$.*

To prove this, we begin by observing that we can write $q$ as a sum of functions each confined to a single coset of $C$; consequently it will suffice to consider the case where $q$ itself vanishes outside the coset $h + C$. We now pause to apply Lemma 1 to the case of the $L_1$-algebra of an infinite cyclic group, generated say by $h'$. For any $\epsilon$, this provides us with an element $z'$ satisfying

$$\| h'z' - z' \| < \epsilon, \tag{1}$$

and such that the Fourier transform of $z'$ is 1 in a neighborhood of $O$. Suppose $z'$ is equal to $a_n$ at the coördinate $nh'$. We return to $L_1(H)$ and there define $z$ to be equal to $a_n$ on the coset $nh + C(n = 0, \pm 1, \ldots)$ and otherwise 0.[5] Then the Fourier transform of $z$ is again 1 in a neighborhood of 0. Moreover if we denote the characteristic function of $h + C$ by $f$, the inequality (1) may be transferred to read $\|fz - z\| < \epsilon$ (we are assuming the measure of $C$ to be 1). It can be verified that $qz = cfz$. Now $(p, q - qz + cz - ce)$ is in $I$, since[6] its Fourier transform vanishes in a neighborhood of 0. As $\epsilon \rightarrow 0$, $qz - cz \rightarrow 0$. Since $I$ is closed, we deduce that $(p, q - ce)$ is in $I$.

We now complete the proof of Theorem 1. Let $m$ denote any element in $L_1(G)$ whose Fourier transform vanishes at 0, that is $\int_G m = 0$. We have to prove that $m$ is in $I$. We may choose $s_i$, $t_i (i = 1, \ldots, r)$ in $L_1(V)$, $L_1(H)$ such that

$$\| m - \sum (s_i, t_i) \| < \epsilon. \tag{2}$$

Write $\int_V s_i = a_i$, $\int_H t_i = b_i$; then

$$\left| \sum a_i b_i \right| < \epsilon. \tag{3}$$

Choose $f_n \epsilon L_1(V)$ as in Lemma 1, with $n$ so large that

$$\| s_i f_n - a_i f_n \| < \epsilon / \sum \| t_i \| \tag{4}$$

for each $i$. By Lemma 2,

$$(a_i f_n, t_i - b_i e) \epsilon I. \tag{5}$$

Finally,

$$(s_i f_n - s_i, t_i) \epsilon I, \tag{6}$$

since its Fourier transform vanishes in a neighborhood of 0. Combining (2)–(6), we find that $m$ differs from an element of $I$ by less than $\epsilon(3 + \| f_n \|)$. This concludes the proof of Theorem 1.

*Remark:* If Lemma 1 were valid in $H$ as well as in $V$, it would be quite simple to extend it to $G$. However, I have been able to do this only on the assumption that $H/C$ is finitely generated. A straightforward argument then leads to the following result.[7]

THEOREM 2. *Suppose, in the above notation, that $H/C$ is finitely generated.*

*Let I be a closed ideal in* $L_1(G)$ *such that the set* $\{M_i\}$ *of regular maximal ideals containing I has a countable frontier. Then* $I = \cap M_i$.

*Applications:* (1) Godement[8] states that, on the basis of Theorem 1, it is possible to extend Beurling's[3] results to any locally compact abelian group.

(2) Let $T$ be an operator in a Banach space such that $\| T^n \| < K$ for all (positive and negative) $n$. We may show that $T$ resembles a normal operator in Hilbert space to the following extent: if $T$ has only 1 in its spectrum, then $T$ is the identity operator.[9] To prove this, one considers the homomorphism of the $L_1$-algebra of the integers given by $\{a_n\} \rightarrow \sum a_n T^n$. The kernel is easily seen to be a closed primary ideal. Its maximality gives the desired result.

(3) The preceding example in effect discussed a representation of the group of integers. More generally, Godement[8] studies representations of any locally compact abelian group by operators in a Banach space; assuming Theorem 1, he proves that the only irreducible ones are one-dimensional. The following more general result is also a corollary of Theorem 1.

THEOREM 3. *Let G be the direct sum of a compact group C and a locally compact abelian group D. Then any irreducible representation of G by operators in a Banach space is finite-dimensional.*

To prove this, we consider the induced representation of $D$ and its extension (S, p. 79) to $L_1(D)$. The kernel of this extension can be seen to be a primary ideal. By Theorem 1, the latter is maximal. Hence the elements of $D$ are represented by scalar operators. This shows that the induced representation of $C$ is also irreducible, from which it follows that the representation is finite-dimensional.[10]

As a corollary of Theorem 3, we note that if $P$ is a primitive ideal in $L_1(G)$, then $L_1(G)/P$ is finite-dimensional and hence regular maximal; this proves Theorem 1.7 of S in a somewhat sharper form. We thus have a natural one-one correspondence between irreducible representations of $G$ and primitive ideals in $L_1(G)$. With this result in hand, we may proceed down the list of the standard theorems known to hold in the abelian and compact cases. In particular, Theorem 1 of this paper extends to the direct sum of an abelian and a compact group.

---

[*] The author is a fellow of the John Simon Guggenheim Memorial Foundation, on leave from the University of Chicago.

[1] *C. R. Acad. Sci. Paris*, 227, 424–426 (1948).

[2] *Uchenye Zapiski Moskov, Gos. Univ. Mat.*, 30, 83–130 (1939).

[3] *Acta Math.*, 77, 127–136 (1945).

[4] *Trans. Am. Math. Soc.*, 61, 69–105 (1947). This paper will be cited as S.

[5] We are considering explicitly only the case where the coset $h + C$ is of infinite order in $H/C$; the case where it is of finite order is considerably simpler.

[6] S, Theorem 2.2.

[7] For the groups of Lemma 1, this extension of Theorem 1 is implicit in the work of Ditkin and Segal, and is mentioned explicitly by Schwartz.

[8] *Ann. École Norm.*, **74**, 119–138 (1946). Theorem 1 is hypothesis A on page 136.

[9] For a reflective Banach space, this theorem can also be deduced from results of Lorch, *Trans. Am. Math. Soc.*, **49**, 18–40 (1941), and a stronger result can also be proved by using a theorem of Gelfand and Hille.

[10] It is well known that any irreducible representation of a compact group by unitary operators in Hilbert space is finite-dimensional. The methods which prove this can be extended to the case of a representation on any Banach space.

---

# *INTEGRAL REPRESENTATIONS OF BILINEAR FUNCTIONALS*[*]

By Marston Morse and William Transue

The Institute for Advanced Study, Princeton, New Jersey

Communicated January 15, 1949

1. *Introduction.*—Let $L_2$ designate the linear space of measurable functionals $x$ with values $x(s)$ defined over the interval $[0, 1]$ and with a pseudo-norm

$$|x|_{L_2} = [\int_0^1 x^2(s)\, ds]^{1/2}.$$

We term $|x|_{L_2}$ a *pseudo-norm* rather than a norm, since we do not find it convenient to require that $x$ be the null element in $L_2$ when $|x|_{L_2} = 0$. The null element in $L_2$ shall instead be the functional with null values. The space $C$ is that defined by Banach (p. 11). Let $A$ be $C$ or $L_2$ and $B$ be $C$ or $L_2$. A functional $F$ is *bilinear* over $A \times B$ by definition if $F(x, y)$ is defined for each $x$ in $A$ and $y$ in $B$, if the functionals $F(x, \cdot)$ and $F(\cdot, y)$, defined by $F$ for fixed $x$ and $y$, respectively, are additive and homogeneous over $B$ and $A$, and if $|F(x, y)| \leqq |x|_A |y|_B M$ for some constant $M$ and every $(x, y)$ in $A \times B$.

*The Space A:* With $A = C$ or $L_2$ we shall define a space $A_d$ (space of distribution functions) associated with $A$. Let $g$ map an axis of reals $R_1$ into $R_1$. Let $g$ be termed *canonical* rel. to $[0, 1]$ if $g(s) = 0$ for $s \leqq 0$ and $g(s) = g(1)$ for $s \geqq 1$, and if $g$ is left continuous except at most at $s = 1$. The space $A_d$ shall be the linear space of functionals $g$ canonical rel. to $[0, 1]$, with a finite total variation $T(g)$ and norm (in the sense of Banach, p. 53) $|g|_{A_d} = T(g)$ when $A = C$; and with a finite Hellinger integral [cf. Hobson, Vol. 1, p. 673] when $A = L_2$ and a norm

$$|g|_{A_d} = \left[\int_0^1 \frac{[dg]^2}{ds}\right]^{1/2} \qquad (1.1)$$

*The Variation h(A, B, k):* We admit partitions $\pi$ of $[0, 1]$ into intervals $[0, s_1)[s_1, s_2) \dots [s_{r(\pi)-1}, 1]$ open at the right when the parenthesis ) is used. An $s$-function (step-function) $\eta$ *associated* with $\pi$ and *belonging* to

[0, 1] by definition maps $R_1$ into $R_1$ with $\eta(s) = 0$ for $s$ not in [0, 1] and $\eta(s) = \eta_r$, a constant on the $r$th interval of $\pi$. Let $(s, t)$ be Cartesian coördinates in a plane $R_2$. Let $E'$ and $E''$ respectively represent the interval [0, 1] on the $s$- and $t$-axes. Let $\pi'$ represent the above partition of $E'$ and let $\pi''$, with vertices $0 = t_0 < t_1 < \ldots < t_{n(\pi'')} = 1$ be a similar partition of $E''$. Let $\eta$ and $\zeta$ be $s$-functions associated respectively with $\pi'$ and $\pi''$. Set

$$\Delta_{rn}(k) = k(s_r, t_n) - k(s_{r-1}, t_n) - k(s_r, t_{n-1}) + k(s_{r-1}, t_{n-1}) \qquad \begin{cases} r = 1, \ldots r(\pi') \\ n = 1, \ldots n(\pi'') \end{cases} \tag{1.2}$$

for any $k$ mapping $R_2$ into $R_1$. Let $A \times B$ be any one of the products $C \times C$, $C \times L_2$, $L_2 \times L_2$. We say that $k$ is canonical rel. $E' \times E''$ if $k(s, \cdot)$ is canonical rel. $E''$ for each fixed $s$, and $k(\cdot, t)$ is canonical rel. to $E'$ for each fixed $t$. With $k$ canonical rel. $E' \times E''$ set

$$h(A, B, k) = \sup_{\eta, \zeta}. \eta_r \zeta_n \Delta_{rn}(k) \tag{1.3}$$

summing as to $r$ and $n$ as in tensor algebra, and taking the sup. over all $\eta$ and $\zeta$ for which $|\eta|_A \leqq 1$, $|\zeta|_B \leqq 1$. We understand thereby that $|\eta|_C = \max. |\eta(s)|$. If $h(A, B, k) < \infty$, $k$ is termed a $d$-function rel. $A \times B$. A principal theorem follows. See MT (2), Theorem 12.1 and Lemma 12.3, MT (3), Theorems 3.3 and 4.3.

THEOREM 1.1. *A functional $\phi$ bilinear over $A \times B$ has a value for $(x, y) \in A \times B$ given by a repeated L-S-integral*

$$F(x, y, k) = \int_0^1 x(s)\, d_s \int_0^1 y(t) d_t k(s, t) = \int^1 y(s) d_s \int_0^1 x(t) d_t k(t, s), \tag{1.4}$$

*where $k$ is a d-function rel. $A \times B$ uniquely determined by $\phi$ and*

$$h(A, B, k) = \sup_{x, y}. \frac{|\phi(x, y)|}{|x|_A |y|_B} \qquad \{|x|_A > 0, |y|_B > 0\}. \tag{1.5}$$

*Conversely if $k$ is a d-function rel. $A \times B$, the integrals (1.4) exist and define a functional $\phi$ bilinear over $A \times B$, for which (1.5) holds.*

We shall need the following, see MT (2), Lemma 5.1 and Theorem 5.2, MT (3), Theorems 3.3 and 4.3.

THEOREM 1.2. *A functional $g$ mapping $R_1$ into $R_1$ and canonical rel. [0, 1] is in $A_d$ if and only if* sup. $|\eta_r \Delta_r(g)| < \infty$ *taking the sup. as in (1.3) [with $\Delta_r(g) = g(s_r) - g(s_{r-1})$]. If $g$ is in $A_d$*

$$\sup_{\eta}. |\eta_r \Delta_r(g)| = |g|_{A_d}. \tag{1.6}$$

Let $p$, $q$, $r$, respectively, be $d$-functions rel. to $C \times C$, $C \times L_2$, $L_2 \times L_2$

with $p = p^*$ and $r = r^*$ where $p^*(s, t) = p(t, s)$, $r^*(s, t) = r(t, s)$. For fixed $s$ and $t$ it has been shown in MT (2), paragraph 7, that

$$|p(s, \cdot)|_{C_d} \leqq h(C, C, p); \qquad |p(\cdot, t)|_{C_d} \leqq h(C, C, p); \qquad |q(\cdot, t)|_{C_d} \leqq h(C, L_2, q) \tag{1.7}$$

$$|q(s, \cdot)|_{L_{2d}} \leqq h(C, L_2, q); \qquad |r(s, \cdot)|_{L_{2d}} \leqq h(L_2, L_2, r); \qquad |r(\cdot, t)|_{L_{2d}} \leqq h(L_2, L_2, r). \tag{1.8}$$

Given $u \in C$ and $y \in L_2$ the $L$-$S$-integrals

$$\varphi_u(s) = \int_0^1 u(t)d_t p(s, t) \qquad \psi_y(s) = \int_0^1 y(t)d_t q(s, t) \tag{1.9}$$

$$f_u(s) = \int_0^1 u(t)d_t q(t, s) \qquad h_y(s) = \int_0^1 y(t)d_t r(s, t) \tag{1.10}$$

define functionals $\varphi_u$ and $\psi_y$ in $C_d$ and functionals $f_u$ and $h_y$ in $L_{2d}$. See MT (2), paragraph 9.

*Objective:* Let $u(s)$ and $v(s)$ map $[0, 1]$ into $R_1$ and be indefinite integrals of functionals in $L_2$. We say that $u$ and $v$ are in $IL_2$. *Necessary and sufficient conditions that the bilinear sum*

$$S(u, v;\ p, q, r) = F(u, v, p) + F(u, \dot{v}, q) + F(v, \dot{u}, q) + F(\dot{u}, \dot{v}, r) = 0$$

*for every* $u$, $v$ *in* $IL_2$ *will be obtained.* It is seen that $S(u, v;\ p, q, r) = S(v, u;\ p, q, r)$. The bilinear sum $S$ includes the classical second variation as a very special case. Its theory will be developed so as to extend Hilbert space theory and the theory of integro-differential equations. The present note contains preliminary theorems of fundamental nature.

2. *Lemma 2.1 and Formula (2.5); Measurability.*—We need the relation

$$h(C, C, k) \leqq h(A, B, k) \qquad [A = C \text{ or } L_2, B = C \text{ or } L_2]. \tag{2.1}$$

It follows from the definition of $h(A, B, k)$ and the relations $|\eta|_A \leqq |\eta|_C$, $|\zeta|_B \leqq |\zeta|_C$ that

$$|\eta_r\zeta_n\Delta_{rn}(k)| \leqq |\eta|_A|\zeta|_B\, h(A, B, k) \leqq |\eta|_C|\zeta|_C\, h(A, B, k). \tag{2.2}$$

On taking the sup. of the extreme members of (2.2) with $|\eta|_C \leqq 1$, $|\zeta|_C \leqq 1$ (2.1) follows. Thus $h(C, C, k) < \infty$ for a $d$-function $k$, and we shall show that $k$ is then measurable. That $k$ is measurable over the complement of the open square $(0, 1) \times (0, 1)$ is immediate, since $k(\cdot, 1)$ and $k(1, \cdot)$ are of B.V. by (1.7), and (1.8). According to Theorem 2.1 of MT (1)

$$\lim_{(s, t) \to (0, 0)} k(a - s, b - t) = k(a, b) \qquad (s \geqq 0, t \geqq 0) \tag{2.3}$$

at each point $(a, b)$ of $[0, 1) \times [0, 1)$. Let $[0, 1) \times [0, 1)$ be divided into a network of squares $Q_{rm}$

$$\frac{r-1}{n} \leqq s < \frac{r}{n} \quad \frac{m-1}{n} \leqq t < \frac{m}{n} \qquad (r, m = 1, \ldots, n).$$

Let $k_n$ map $[0, 1) \times [0, 1)$ into $R_1$, with $k_n(s, t)$ constant on $Q_{rm}$ and $k(s, t) = k_n(s, t)$ when $s = \frac{r-1}{n}$, $t = \frac{m-1}{n}$. It is trivial that $k_n$ is measurable. It follows from (2.3) that $k_n(s, t) \to k(s, t)$ over $[0, 1) \times [0, 1)$ as $n \to \infty$, so that $k$ is measurable.

LEMMA 2.1. *With $q$ defined as above the integral*

$$\mu(s, t) = \int_0^s q(a, t)da \qquad [(s, t) \in E' \times E''] \tag{2.4}$$

*properly extended over the $(s, t)$-plane defines a d-function $\mu$ rel. to $L_2 \times L_2$ with $h(L_2, L_2, \mu) \leqq h(C, L_2, q)$.*

Since $q(\cdot, t)$ is of B.V. for fixed $t$ by (1.7), the integral (2.4) exists. For fixed $s$, $q(s, \cdot)$ is absolutely continuous over $E''$ by (1.8), since its Hellinger integral is finite. Hence $\mu(s, t)$ is continuous over $E''$ for fixed $s$. It is obvious that $\mu(\cdot, t)$ is continuous over $E'$ for fixed $t$. Moreover $\mu(0, t) = \mu(s, 0) = 0$ for $t \in E''$ and $s \in E'$. So defined over $E' \times E''$, $\mu$ can be extended in a unique manner over $R_2$ so as to be canonical rel. $E' \times E''$. See MT (2), Lemma 6.1. It remains to show that $h(L_2, L_2, \mu) \leqq h(C, L_2, q)$. To that end let $\eta$ and $\zeta$ be $s$-functions belonging to $E'$ and $E''$, respectively, with $|\eta|_{L_2} \leqq 1$, $|\zeta|_{L_2} \leqq 1$. For fixed $s \in E''$ set $\zeta_n \Delta_n[q(s, \cdot)] = \lambda(s)$, $\zeta_n \Delta_n[\mu(s, \cdot)] = \tau(s)$. Then $\tau$ is a. c., and $\dot{\tau}(s) = \lambda(s)$ p.p. (presque partout) over $E'$. Moreover $\lambda$ is canonical rel. $E'$, since $q(\cdot, t)$ is so canonical for fixed $t$. It follows from Theorem 1.2 that $|\lambda(s)| \leqq |q(s, \cdot)|_{L_{2d}}$ and so from (1.8) that $|\lambda(s)| \leqq h(C, L_2, q)$. Then

$$|\eta_r \zeta_n \Delta_{rn}(\mu)| = |\eta_r \Delta_r(\tau)| \leqq \left| \int_0^1 \eta(s)\lambda(s)\, ds \right| \leqq h(C, L_2, q)$$

using the Schwarz lemma. The concluding relation of the lemma follows.

Given a measurable $d$-function $k$ set

$$k_1(s, t) = \limsup_{n \to \infty} \left[k\left(s + \frac{1}{n}, t\right) - k(s, t)\right]n \qquad (n = 1, 2, \ldots).$$

Then $k_1$ is measurable and equals the $s$-derivative of $k$ where it exists. One defines $k_2$ similarly with reference to $t$. Since $p$, $q$, $r$ have been shown to be measurable, $p_1$, $p_2$, etc., are measurable. We shall establish the basic formula.

$$\int_0^s da \int^1 \dot{u}(t)d_t q(a, t) = \int^1 \dot{u}(t)d_t \int_0^s q(a, t)da = \int_0^1 \dot{u}(t)d_t\mu(s, t) \qquad [u \in IL_2]. \tag{2.5}$$

Note first that

$$\int_0^1 |q_2(s, t)|dt \leqq |q_2(s, \cdot)|_{L_2} = |q(s, \cdot)|_{L_{2d}} \leqq h(C, L_2, q) \tag{2.6}$$

using a known property of Hellinger integrals and (1.8). Since $q(s, \cdot)$ is a. c. and (2.6) holds

$$\mu(s, t) = \int_0^s d\beta \int_0^t q_2(\beta, \alpha)d\alpha = \int_0^t d\alpha \int_0^s q_2(\beta, \alpha)d\beta \tag{2.7}$$

using Tonelli's Theorem. [See Fubini, Tonelli.] The left member of (2.5) becomes

$$\int_0^s da \int_0^1 \dot{u}(t)q_2(a, t)dt = \int_0^1 \dot{u}(t)\left[\int_0^s q_2(a, t)da\right]dt = \int_0^1 \dot{u}(t)\mu_2(s, t)dt$$

using Tonelli's Theorem and (2.7). Relation (2.5) follows on noting that $\mu(s, \cdot)$ is a. c. over [0, 1] for fixed $s$.

3. *Special d-Functions.*—Certain special $d$-functions will be needed to decompose $S$. A *triangle* $d$-function $k$ shall be such that $k = k^*$ and $k(s, t) = k(s, 1)$ for $0 \leqq s \leqq t \leqq 1$. For an *edge* $d$-function $k(s, t) = 0$ for $s < 1$; as canonical rel. $E' \times E''$, $k$ is determined by the values of $k(1, t)$ for $0 \leqq t \leqq 1$. For a *vertex* $d$-function $k(s, t) = 0$ on $E' \times E''$ except at most at (1, 1); $k$ is then determined by $k(1, 1)$. Each such function is required to be canonical rel. $E' \times E''$. Given $p$, $q$, $r$ as previously set

$$\begin{aligned} r_p(s, t) &= \int_0^s \int_0^t p(\alpha, \beta)d\alpha\, d\beta \qquad [(s, t) \in E' \times E''] \\ r^{(q)}(s, t) &= \int_0^s q(a, t)da + \int_0^t q(a, s)da. \end{aligned}$$

Let $\tau_p$ be the triangle function with $\tau_p(s, 1) = p(s, 1)$ on $E'$.

Let $\tau^{(p)}$ be the triangle function with $\tau^{(p)}(s, 1) = \int_0^s p(a, 1)da$ on $E'$.

Let $e^{(q)}$ be the edge function with $e^{(q)}(1, t) = q(1, t)$ on $E''$.

Let $\nu_p$ be the vertex function with $\nu_p(1, 1) = p(1, 1)$.

We shall establish the relation

$$\int_0^1 u(s)d_s\int_0^1 v(t)d_t\tau_p(s, t) = u(1)v(1)p(1, 1) - \int_0^1 p(s, 1)d[u(s)v(s)] \quad (3.1)$$

for $u$, $v$ in $IL_2$. On integrating by parts, see Saks, p. 102.

$$\begin{aligned} &\int^1 v(t)d_t\tau_p(s, t) = \int_0^s v(t)dp(1, t) = v(s)p(1, s) - \int_0^s \dot{v}(t)p(1, t)dt \\ &\int_0^1 u(s)d_s \int_0^1 v(t)d_t\tau_p(s, t) = \int_0^1 u(s)d[v(s)p(1, s)] - \\ &\int_0^1 u(s)\dot{v}(s)\, p(1, s)ds = u(1)v(1)p(1, 1) - \int_0^1 \dot{u}(s)v(s)p(1, s)ds - \\ &\qquad\qquad \int_0^1 u(s)\dot{v}(s)p(1, s)ds \quad (3.2) \end{aligned}$$

from which (3.1) follows.

We note the relations, valid for $u, v \in IL_2$

$$\int_0^1 u(s)d_s\int_0^1 v(t)d_t\tau_p(s, t) = \int_0^1 u(s)v(s)dp(s, 1) \quad (3.3)$$

$$\int_0^1 u(s)d_s\int_0^1 \dot{v}(t)d_t\tau^{(p)}(s, t) = \int_0^1 u(s)\dot{v}(s)p(s, 1)\, ds \quad (3.4)$$

$$\int_0^1 u(s)d_s\int_0^1 \dot{v}(t)d_te^{(q)}(s, t) = u(1) \int_0^1 \dot{v}(t)dq(1, t) \quad (3.5)$$

$$\int_0^1 u(s)d_s\int_0^1 v(t)d_t\nu_p(s, t) = u(1)v(1)p(1, 1) \quad (3.6)$$

first verifying the existence of the integrals involved.

Relation (3.3) follows from (3.1) on integration by parts, and (3.4)–(3.6) are immediate.

LEMMA 3.1. *The functionals* $r_p$ *and* $r^{(q)}$ *are d-functions rel.* $L_2 \times L_2$,

*$\tau^{(p)}$ and $e^{(q)}$ are d-functions rel. $C \times L_2$, while $\tau_p$ and $\nu_p$ are d-functions rel. $C \times C$.*

That $r^{(q)}$ is a $d$-function rel. $L_2 \times L_2$ follows from Lemma 2.1 on noting that $r^{(q)} = \mu + \mu^*$. Continuing

$$\eta_r\zeta_n\Delta_{rn}(r_p) = \int_0^1 \int_0^1 \eta(s)\zeta(t)p(s, t)\, ds\, dt$$

from which it follows that

$$h(L_2, L_2, r_p) \leqq \max. \, |p(s, t)| \leqq h(C, C, p).$$

Given a $d$-function, $k(s, t)$ and a partition $\pi''$ of $E''$ we denote by $\Delta_n^{(2)}(k)$ the functional with values $\Delta_n^{(2)}[k(s, \cdot)] = k(s, t_n) - k(s, t_{n-1})$. If $\eta$ is an $s$-function and $g \in A_d$ then $\int_0^1 \eta(s)dg(s) = \eta_r\Delta_r g$. Then for $k(s, \cdot) \in B_d$ and $\zeta$ an $s$-function

$$\eta_r\zeta_n\Delta_{rn}(k) = \eta_r\Delta_r[\zeta_n\Delta_n^{(2)}(k)] = \eta_r\Delta_r\int_0^1 \zeta(t)d_tk(s, t). \tag{3.7}$$

In finding the sup. of the sum (3.7) we can suppose $\pi' = \pi''$ without loss of generality. With this supposed, using (3.7)

$$|\eta_r\zeta_n\Delta_{rn}(\tau_p)| = |\eta_r\Delta_r\int_0^s \zeta(t)d_tp(1, t)| = |\eta_r\zeta_r\Delta_r[p(1, \cdot)]| =$$
$$|\int_0^1 \eta(s)\zeta\ (s)\, dp(1, s)|$$

$$|\eta_r\zeta_n\Delta_{rn}(\tau^{(p)})| = |\int_0^1 \eta(s)\zeta(s)p(1, s)ds|$$

$$|\eta_r\zeta_n\Delta_{rn}(e^{(q)})| = |\eta(1) \int_0^1 \zeta(t)dq(1, t)|$$

from which it follows that

$$h(C, C, \tau_p) \leqq T[p(1, \cdot)] \leqq h(C, C, p) \tag*{(3.3)$'$}$$

$$h(C, L_2, \tau^{(p)}) \leqq |p(1, \cdot)|_{L_2} \leqq h(C, C, p) \tag*{(3.4)$'$}$$

$$h(C, L_2, e^{(q)}) \leqq |q_2(1, \cdot)|_{L_2} \leqq h(C, L_2, q). \tag*{(3.5)$'$}$$

Finally $h(C, C, \nu_p) \leqq |p(1, 1)| \leqq h(C, C, p)$ and the proof of the lemma is complete.

The fact that the special functions $r_p$, $r^{(p)}$, etc., are $d$-functions permits us to draw upon the properties (1.7) to (1.10) and to invert the order of integration in (3.3) to (3.6).

4. *A Model Form for S.*—Three formulae are needed.

$$F(u, v, \tau_p) = F(u, v, \nu_p) - F(u, \dot{v}, \tau^{(p)}) - F(v, \dot{u}, \tau^{(p)}) \tag{4.1}$$

$$F(u, v, p) = F(\dot{u}, \dot{v}, r_p) \qquad [\text{if } p(s, 1) = 0 \text{ on } E'] \tag{4.2}$$

$$F(\dot{u}, \dot{v}, r^{(q)}) = -F(u, \dot{v}, q) - F(v, \dot{u}, q) \qquad [\text{if } q(1, t) = 0 \text{ on } E'']. \tag{4.3}$$

Formula (4.1) is a consequence of (3.1), using (3.6) and (3.4), while (4.2) results on performing both integrations in $F(u, v, p)$ by parts. To verify (4.3) perform the first integration in $F(u, \dot{v}, q)$ by parts. Then

$$F(u, \dot{v}, q) = -\int_0^1 \dot{u}(s)[\int_0^1 \dot{v}(t)d_tq(s, t)]\, ds = \qquad (4.4)$$

$$-\int_0^1 \dot{u}(s)d_s \int_0^s da \int_0^1 \dot{v}\,(t)\, d_t\, q(a, t)$$

$$= -\int_0^1 \dot{u}(s)d_s \int_0^1 \dot{v}(t)d_t\mu(s, t) \qquad [\text{using } (2.5)] \qquad (4.5)$$

$$F(v, \dot{u}, q) = -\int_0^1 \dot{v}(s)d_s \int_0^1 \dot{u}(t)d_t\mu(s, t) \qquad (\text{from } (4.5))$$

$$= -\int_0^1 \dot{u}(s)d_s\int_0^1 \dot{v}(t)d_t\mu(t, s) \qquad (4.6)$$

since $\mu$ is a $d$-function rel. to $L_2 \times L_2$. Formula (4.3) is a consequence of (4.5) and (4.6) and the definitions of $\mu$ and $r^{(q)}$.

DEFINITION. *Two sets of admissible d-functions* $(p, q, r)$ *and* $(p', q', r')$ *are termed equivalent if* $S(u, v;\ p, q, r) = S(u, v;\ p', q', r')$ *for every* $(u, v)$ *in* $IL_2$.

This equivalence relation is transitive, reflexive and symmetric. Two equivalences $(p, q, r) \sim (p', q', r')$ and $(p'', q'', r'') \sim (p''', q''', r''')$ imply the equivalence $(p + p'', q + q'', r + r'') \sim (p' + p''', q' + q''', r' + r''')$. Moreover $(p, q, r) \sim (p', q', r')$ implies $(cP, cq, cr) \sim (cp', cq', cr')$ for every constant $c$.

THEOREM 4.1. *The d-functions* $(p, q, r)$ *are equivalent to a unique set* $(\bar{p}, \bar{q}, \bar{r})$ *of d-functions in which* $\bar{q}$ *is an edge d-function and* $\bar{p}$ *a vertex d-function.*

Since $p - \tau_p$ vanishes on the edges of $E' \times E''$ (4.2) implies that $(p - \tau_p, 0, 0) \sim (0, 0, r_p - r_{\tau_p})$ while (4.1) implies that $(\tau_p, 0, 0) \sim (\nu_p, -\tau^{(p)}, 0)$. On adding the trivial equivalence $(0, q, r) \sim (0, q, r)$ to these equivalences $(p, q, r) \sim (\nu_p, q', r')$ for admissible $q'$ and $r'$. There thus exist $d$-functions $(p', q', r') \sim (p, q, r)$ with $p'$ a vertex $d$-function. From (4.3)

$$(0, q' - e_{q'}, 0) \sim (0, 0, -r^{(q')} + r^{(eq')})$$

since $q' - e_{q'}$ vanishes on $E''$. It follows that $(p', q', r') \sim (p', e_{q'}, \bar{r})$ for an admissible $\bar{r}$, thereby establishing the existence of a *model* set $(\bar{p}, \bar{q}, \bar{r})$ of the required type.

It follows from (3.5) and (3.6) that

$$S(u, v;\ p, q, r) = F(\dot{u}, \dot{v}, \bar{r}) + u(1)v(1)\bar{p}\,(1,1) + u(1)\int_0^1 \dot{v}(t)d_t\bar{q}(1, t) +$$

$$v(1)\int_0^1 \dot{u}(t)d_t\bar{q}(1, t). \qquad (4.7)$$

To show that the *model* set $(p, \bar{q}, \bar{r})$ is uniquely determined by $S$ let $i$ represent a functional in $IL_2$ with $i(s) \equiv 1$.

Let $u_a(0 \leqq a < 1)$ be a functional in $IL_2$ such that $u(1) = 0$ and $\dot{}_a$ is p.p. the characteristic function of $[0, a]$. Then $u_a(s) = s - a$ for $0 \leqq s \leqq a < 1$ while $u_a(s) = 0$ for $a \leqq s \leqq 1$. It follows from (4.7) that

$$\bar{p}(1, 1) = S(i, i;\ p, q, r) \qquad (4.8)'$$

$$\bar{q}(1, a) = S(i, u_a;\ p, q, r) \qquad (0 \leqq a < 1) \qquad (4.8)''$$

$$\bar{r}(a, b) = S(u_a, u_b;\ p, q, r) \qquad (0 \leqq b < 1). \qquad (4.8)'''$$

The remaining values $\bar{q}(1, 1)$, $\bar{r}(a, 1)$ and $\bar{r}(1, b)$ are determined by the values given in (4.8), taking account of the continuity of $\bar{q}(1, \cdot)$ and $\bar{r}(1, \cdot)$. Thus the model set $(\bar{p}, \bar{q}, \bar{r})$ is uniquely determined by $S$. This completes the proof of the theorem.

In terms of $(p, q, r)$ the corresponding model set $(\bar{p}, \bar{q}, \bar{r})$ may be given as follows for $(a, b) \in E' \times E''$ using (4.8).

$$\bar{p}(1, 1) = p(1, 1) \qquad (4.9)'$$

$$\bar{q}(1, a) = q(1, a) + \int_0^a (t - a)dp(1, t) = q(1, a) - \int_0^a p(1, t)dt \quad (4.9)''$$

$$\bar{r}(a, b) = r(a, b) - \int_0^a q(s, b)\, ds - \int_0^b q(s, a)\, ds + \int_0^a \int_0^b p(s, t)\, ds\, dt. \qquad (4.9)'''$$

Corollary 4.1. *A necessary and sufficient condition that $S(u, v;\ p, q, r) = 0$ for every $u, v$ in $IL_2$ is that the right members of (4.9) vanish for every $(a, b)$ in $E' \times E''$.*

---

* The work of Dr. Transue on this paper was done with the financial aid of the Office of Naval Research.

Banach, S., *Théorie des Opérations Linéaires*, Warsaw, 1932.

Fubini, G., "Sugli integrali multipli," *Atti accad. naz. Lincei, Rend.*, **16** (1907).

Hobson, E. W., *The Theory of Functions of a Real Variable*, Vol. I, Cambridge Univ. Press, 1927.

Morse, M., and Transue, W: (1) "Functionals of Bounded Fréchet Variation," *Can. J. Math.*, **I** (1949); (2) "Functionals *F* Bilinear Over the Product $A \times B$ of Two $p$-Normed Vector Spaces. I. The Representation of *F*," *Ann. Math.*, **50** (1949); (3) "Functionals *F* Bilinear Over the Product $A \times B$ of Two $p$-Normed Vector Spaces. II. Admissible Spaces *A*" (to be published).

Saks, S., *Theory of the Integral*, Warsaw-Lwow, 1937.

Tonelli, L., "Sull'integrazione per parti," *Atti accad. naz. Lincei*, **5**, 18 (1909).

---

# *THE COEFFICIENTS OF SCHLICHT FUNCTIONS, IV**

By A. C. Schaeffer and D. C. Spencer

Department of Mathematics, Purdue University, and Department of Mathematics, Stanford University

Communicated by S. Lefschetz, December 17, 1948

1. In the preceding note (reference 2) a method was outlined for attacking the coefficient problem of schlicht functions, but lack of space made it necessary to omit certain lines of attack which have proved useful

in investigating the problem. Moreover, during the last two years new results have been obtained. It is the purpose of this note to give a brief sketch of the present status of the authors' research on the coefficient problem. Although some of the results discussed below antedate note 2, they have not been published elsewhere.

Let $\mathcal{S}$ denote the family of all functions

$$w = f(z) = z + a_2z^2 + a_3z^3 + \ldots \tag{1.1}$$

which are regular and schlicht in the circle $|z| < 1$, and let $V_n$ denote the region in euclidean space of $2n - 2$ real dimensions which is composed of all points $(a_2, a_3, \ldots, a_n)$ belonging to functions $f$ of $\mathcal{S}$. The region $V_n$ is closed, and it is topologically equivalent to the closed $(2n - 2)$-dimensional full sphere. The coefficient problem is the problem of determining $V_n$ for $n = 2, 3, \ldots$.

It was pointed out in reference 2 that any function $w = f(z)$ belonging to a boundary point of $V_n$ satisfies a differential equation of the form

$$\left(\frac{z}{w}\frac{dw}{dz}\right)^2 P(w) = Q(z), \tag{1.2}$$

where

$$P(w) = \sum_{\nu=1}^{n-1} \frac{A_\nu}{w^\nu} \qquad Q(z) = \sum_{\nu=-(n-1)}^{n-1} \frac{B_\nu}{z^\nu}, \qquad B_{-\nu} = \bar{B}_\nu. \tag{1.3}$$

The function $Q(z)$ is non-negative on $|z| = 1$ and has at least one zero there. There is a one-one correspondence between boundary points of $V_n$ and functions $w = f(z)$ of class $\mathcal{S}$ which satisfy a differential equation of the form (1.2). Functions of class $\mathcal{S}$ which satisfy a differential equation of the form (1.2) will be called $\mathfrak{D}_n$-functions.

Starting from the finite zeros of $P(w)$ and from $w = \infty$, prolong all possible loci

$$Re \int \sqrt{P(w)}\,\frac{dw}{w} = \text{constant}.$$

The set of these loci will be denoted by $\Gamma_w$. It can be shown that $\Gamma_w$ dissects the plane into finitely many simply-connected domains $O_\nu$ in each of which the hyperelliptic integral

$$\zeta = \int \sqrt{P(w)}\,\frac{dw}{w}$$

is regular and defines a schlicht mapping of the domain onto a half-plane or a strip. The structure of $\Gamma_w$ formed the underlying basis of the method and results described in reference 2.

2. Let $2m$ arcs ($m \geq 1$) of the unit circumference $|z| = 1$ be identified in pairs such that: (i) the circle $|z| \leq 1$ with boundary arcs identified in this way is a closed surface of genus zero (therefore topologically equivalent to the sphere); (ii) if two identified arcs abut, the common end-point is a zero of $Q$ with at most one exception; (iii) in the identification of any pair of arcs the metric defined by

$$|dZ|^2 = \left|\frac{Q(z)}{z^2}\right| |dz|^2$$

is preserved. Two points which are identified with each other may be said to be equivalent. There are the following classes of equivalent points: an interior point of $|z| < 1$ which is equivalent only to itself; an interior point of an arc on $|z| = 1$ which is equivalent to exactly one other point; an end-point of an arc (vertex) which may be equivalent to several other vertices on $|z| = 1$. Let $z_1, z_2, \ldots, z_k$, $k \geq 1$, be a class of equivalent points on $|z| = 1$, and let $2\sigma_\nu$, $\sigma_\nu \geq 0$, be the order of the zero of $Q(z)$ at $z_\nu$. Write

$$\sigma = \sum_{\nu = 1}^{k} (\sigma_\nu + 1),$$

and let $N_1, N_2, \ldots, N_k$, be half-neighborhoods interior to $|z| < 1$ at the points $z_1, z_2, \ldots, z_k$, respectively. In $N_\nu$ let

$$Z_\nu(z) = \int_{z_\nu}^{z} \sqrt{Q(t)}\, \frac{dt}{t} \tag{2.1}$$

and define

$$\tau = \{Z_\nu(z)\}^{2/\sigma}, \qquad z \subset N_\nu. \tag{2.2}$$

If the branches of the functions (2.1) are suitably chosen, the images of the half-neighborhoods $N_\nu$ will fit together in the $\tau$-plane and will form a complete neighborhood of $\tau = 0$. In this manner we define local uniformizers at each point of the closed surface which becomes, therefore, a Riemann surface of genus zero.

Of the classes of equivalent points on $|z| = 1$ we distinguish one and denote it by $\Sigma$. We remark that if $\sigma = 1$ for a class of equivalent points $z_1, z_2, \ldots, z_k$ on $|z| = 1$, then $k = 1$. Moreover, by (ii) there is at most one point on $|z| = 1$ for which $\sigma = 1$. We suppose that $\Sigma$ satisfies the following condition: (iv) $\Sigma$ may be any class of identified points on $|z| = 1$ subject only to the restriction that if $\sigma = 1$ for some point on $|z| = 1$, then $\Sigma$ is that point. Two functions $Q$ will be considered equal if their ratio is a positive constant. Then $Q$ plus an identification of arcs on $|z| = 1$ and a choice of $\Sigma$ satisfying (i), (ii), (iii) and (iv) define a point in a space $S_n$. Each point of $S_n$ defines a Riemann surface having a distinguished point corresponding to $\Sigma$. We have the following result:

*To each point* $\mathfrak{p}$ *of* $S_n$ *there belongs a unique* $\mathfrak{D}_n$*-function* $w = f(z)$ *which maps the closed Riemann surface belonging to* $\mathfrak{p}$ *onto the closed w-plane and carries* $\Sigma$ *into* $w = \infty$. *It satisfies the differential equation*

$$\left(\frac{z}{w}\frac{dw}{dz}\right)^2 P(w) = Q(z), \tag{2.3}$$

*where* $Q$ *is given by the point of* $S_n$ *and*

$$P(w) = \frac{A_1}{w} + \frac{A_2}{w^2} + \ldots + \frac{A_{n-1}}{w^{n-1}}.$$

*To each* $\mathfrak{D}_n$*-function there corresponds a point of* $S_n$ *and, if* $f(z)$ *is not algebraic, it belongs to a unique point of* $S_n$.

We say that the function $w = f(z)$ realizes the identification of arcs on $|z| = 1$ satisfying the conditions (i), (ii) and (iii). The given identification may also be realized in a stepwise fashion or continuously. For example, given an identification satisfying (i), (ii) and (iii) and a distinguished class or cycle $\Sigma$, let $I_1', I_1'', I_2', I_2'', \ldots, I_m', I_m''$ be the pairs of identified arcs on $|z| = 1$. By the definition of $\Sigma$ it is clear that there is always one pair of abutting identified arcs, say $I_1'$, $I_1''$, with a common end-point $e^{-i\alpha}$ not in $\Sigma$. It then follows that $Q(z)$ must be zero at $z = e^{-i\alpha}$. Let $I'$ and $I''$ be identified subarcs of $I_1'$ and $I_1''$ abutting at $e^{-i\alpha}$. Then there is a function

$$v(z, u) = u\{z + b_2z^2 + b_3z^3 + \ldots\}, \; 0 < u < 1, \tag{2.4}$$

which identifies $I'$ and $I''$. The function $v(z, u)$ maps $|z| < 1$ onto $|v| < 1$ minus a piecewise analytic arc whose two edges correspond to $I'$ and $I''$, respectively. If $m > 1$, the arcs $I'$ and $I''$ may be allowed to increase continuously until they coincide with $I_1'$ and $I_1''$. This defines a function $v(z, u)$ in the interval $u_1 \leq u \leq 1$, $u_1 > 0$. The function $v_1 = v(z, u_1)$ maps $I_1'$ and $I_1''$ into the opposite edges of a slit interior to the unit circle and it maps $I_2', I_2'', \ldots, I_m', I_m''$ into a new set of identified arcs $J_2'$, $J_2'', \ldots, J_m', J_m''$ lying on the unit circumference in the plane of $v_1$. The identification of arcs on $|v_1| = 1$ satisfies the same postulates, and we may repeat the process. The following theorem is true:

*Let* $w = f(z)$ *be a* $\mathfrak{D}_n$*-function belonging to a given point of* $S_n$ *and satisfying*

$$\left(\frac{z}{w}\frac{dw}{dz}\right)^2 P(w) = Q(z). \tag{2.5}$$

*Then there is a function*

$$v(z, u) = u\{z + b_2(u)z^2 + b_3(u)z^3 + \ldots\}, \qquad 0 < u \leq 1,$$

*which is regular and schlicht in* $|z| < 1$, *maps* $|z| < 1$ *onto* $|v| < 1$ *minus a finite set of analytic arcs, and satisfies*

$$\left(\frac{z}{v}\frac{\partial v}{\partial z}\right)^2 Q(v, u) = Q(z), \tag{2.6}$$

*where*

$$Q(v, u) = \sum_{\nu = -(n-1)}^{n-1} \frac{B_\nu(u)}{v^\nu}, \qquad B_{-\nu}(u) = \overline{B_\nu(u)},$$

*is non-negative on* $|v| = 1$ *with at least one zero there. Moreover,* $v(z, u)$ *is continuous with respect to* $u$ *and, except at finitely many points, is differentiable and satisfies*

$$u\frac{\partial v}{\partial u} = v\frac{1 + e^{i\alpha}v}{1 - e^{i\alpha}v}, \tag{2.7}$$

*where* $\alpha = \alpha(u)$ *is a piecewise continuous function of* $u$ *and* $Q(e^{-i\alpha(u)}, u) = 0$. *The terminal conditions are*

$$v(z, 1) = z, \qquad \lim_{u \to 0} \frac{v(z, u)}{u} = f(z),$$

$$Q(z, 1) = Q(z), \qquad \lim_{u \to 0} Q(uw, u) = P(w).$$

We remark that (2.7) is the differential equation derived by Löwner for functions mapping $|z| < 1$ onto $|v| < 1$ minus any Jordan arc not passing through the origin. However, the equation (2.6) for the special class of functions under consideration is an automatic consequence of the continuous realization of the identification.

As $u$ varies from 1 to 0, the point $(b_2(u), b_3(u), \ldots, b_n(u))$ traces a curve which extends from the origin $b_2 = 0, b_3 = 0, \ldots, b_n = 0$ and tends to the boundary point of $V_n$ defined by $w = f(z)$ as $u$ tends to 0. This curve lies entirely in the interior of $V_n$ except for its end-point at $u = 0$. Curves such as this have been used in a numerical calculation concerning the boundary of $V_4$ which has been carried out since the winter of 1946–1947 under a project sponsored by the Office of Naval Research. The computation is not yet complete, and we shall therefore give only a brief sketch of it.

One purpose of the computation is to prove that $|a_4| \leq 4$ for functions $f(z)$ of class $\mathcal{S}$. This part of the program consists of three parts: ($a$) proof that points of the boundary of $V_4$ corresponding to functions of the form

$$f(z) = \frac{z}{(1 - e^{i\theta}z)^2} = z + 2^{i\theta}z^2 + 3e^{2i\theta}z^3 + 4e^{3i\theta}z^4 + \ldots \tag{2.8}$$

have the local maximum property with respect to neighboring points; (*b*) proof that over a large part of the boundary $|a_4| \leq 4 - \delta$, where $\delta$ is some positive quantity; (*c*) calculation of points on the remainder of the boundary by integrating a system of differential equations obtained from (2.6) and (2.7).

Equating like powers of $z$ in the equation (2.7) we obtain differential equations for the coefficients $b_\nu(u)$ of the function $v(z, u)$. By integration we then obtain Löwner's integral formulas for the coefficients $b_\nu(u)$. From the formula for $b_4(0) = a_4$ it can be shown that if there is a constant $c$ such that $|\alpha(u) - c| \leq 0.08697$, $0 \leq u \leq 1$, then $|a_4| \leq 4$ with equality if and only if $\alpha(u)$ is almost everywhere equal to a constant. This establishes the desired local maximum property of the functions (2.8).

If $f(z)$ is a function of class $\mathcal{S}$ which maximizes $Re(a_4)$, then it satisfies an equation (1.2) where the coefficients $A_\nu$, $B_\nu$ in (1.3) are of the form

$$\left\{\begin{array}{llll} & A_1 = a_2^2 + 2a_3, & A_2 = 3a_2, & A_3 = 1 \\ B_0 = 3a_4, & B_1 = 3a_3, & B_2 = 2a_2, & B_3 = 1. \end{array}\right.$$

Moreover, $Q(z)$ is of the form

$$Q(z) = \frac{1}{z^3}\,(z + e^{-i(\varphi + \psi)})^2(z^2 - 2pze^{i\varphi} + e^{2i\varphi})(z^2 - 2qze^{i\psi} + e^{2i\psi}),$$

where $p \geq 1$, $q \geq 1$. Comparing the like powers of $z$ we find that

$$a_2 = e^{i(\varphi + \psi)} - qe^{-i\psi} - pe^{-i\varphi}$$

$$3a_3 = e^{2i(\varphi + \psi)} - 4pe^{i\varphi} - 4qe^{i\psi} + e^{-2i\varphi} + e^{-2i\psi} + 4pqe^{-i(\varphi + \psi)}.$$

We add the condition that $|a_2| \leq 2$ and substitute these expressions in the Prawitz form of the area-principle (see reference 1). It can be shown that if $\gamma_1$ and $\gamma_2$ are numbers satisfying $1 < \gamma_1 \leq 1.1$, $1.25 \leq \gamma_2 < \infty$, then the only possible maxima of $a_4$ in the region where $\min(p, q) \leq \gamma_1$ or $\max(p, q) \geq \gamma_2$ at which $a_4 > 3.6$ occur for $\min(p, q) = \gamma_1$ or $\max(p, q) = \gamma_2$. Hence, to exclude the region $\min(p, q) \leq \gamma_1$, $\max(p, q) \geq \gamma_2$ from consideration it is sufficient to show that $a_4 < 4$ when $p = \gamma_1$, $\gamma_1 \leq q \leq \gamma_2$ and when $p = \gamma_2$, $\gamma_1 \leq q \leq \gamma_2$. Suppose then that

$$\gamma_1 \leq p \leq q \leq \gamma_2. \tag{2.9}$$

From considerations of symmetry we may also suppose that $0° \leq \varphi \leq 60°$.

The final step in the calculation is to compute points on the boundary of $V_4$ corresponding to points of $S_4$ satisfying (2.9) (and also $0° \leq \varphi \leq 60°$). If $re^{i\varphi}$, $\rho e^{i\psi}$ are the roots in $|v| < 1$ of the function $Q(v, u)$ in (2.6), we obtain from (2.7) the four equations

$$\left\{\begin{aligned} u\frac{dr}{du} &= r\frac{1-r^2}{1-2r\cos(\alpha+\varphi)+r^2}, & u\frac{d\rho}{du} &= \rho\frac{1-\rho^2}{1-2\rho\cos(\alpha+\psi)+\rho^2}, \\ u\frac{d\varphi}{du} &= 2r\frac{\sin(\alpha+\varphi)}{1-2r\cos(\alpha+\varphi)+r^2}, & u\frac{d\psi}{du} &= 2\rho\frac{\sin(\alpha+\psi)}{1-2\rho\cos(\alpha+\psi)+\rho^2}. \end{aligned}\right.$$

We observe that $\alpha(u) = \varphi + \psi - \pi$. If the starting values satisfy (2.9) (as we suppose), then it can be shown that

$$\left|\frac{d\alpha}{du}\right| \leq \frac{2}{\gamma_1 - 1}. \tag{2.10}$$

We begin at $u = 1$ with values corresponding to the range (2.9) and whenever $\alpha(u)$ deviates from a constant by an amount not exceeding 0.08697, we know that $|a_4| \leq 4$. Since values of $\alpha(u)$ will be obtained at a discrete set of points, we use (2.10) as a means of estimating intermediate values.

3. We conclude with some results concerning the boundary of the general region $V_n$. A point on the boundary of $V_n$ at which there is a $(2n - 3)$-dimensional hyperplane such that $V_n$ lies entirely on one side of it will be called a supporting point. The set of supporting points will be called the supporting surface and it will be denoted by $K_n$. It can be shown that $K_n$ is connected and that any function belonging to a point of $K_n$ maps $|z| < 1$ onto the plane minus a single analytic slit without critical points.

Now let $M_n$ be the set of boundary points of $V_n$ whose points belong to functions $f(z)$ which satisfy one and only one equation (1.2) and in this equation the function $Q(z)$ vanishes at only one point of $|z| = 1$ where it has a zero of precise order two. We have the following theorem:

*The set $M_n$ is not empty, it is open, and its closure contains the supporting surface $K_n$. There is a one-one bicontinuous correspondence between points of $M_n$ and a class of Q-functions vanishing at only one point of $|z| = 1$ where there is a zero of precise order two. If $\mathbf{a}$ is any point of $M_n$ and $f(z)$ the function belonging to $\mathbf{a}$, let $Q(z)$ be the corresponding Q-function and let $e^{-i\alpha}$ be the point on $|z| = 1$ where $Q(z)$ vanishes. Let*

$$\psi(z) = \sum_{\nu=-(n-1)}^{n-1} \frac{\beta_\nu}{z_\nu}, \quad \beta_{-\nu} = \bar{\beta}_\nu,$$

*have a zero of first or higher order at $z = e^{-i\alpha}$, and let*

$$\chi(z) = \frac{1}{2}\int_{e^{-i\alpha}}^{z} \frac{\psi(v)}{\sqrt{Q(v)}}\,\frac{dv}{v},$$

*where* $\sqrt{Q(v)}$ *is analytic on* $|v| = 1$ *except for a discontinuity at the point* $e^{-i\beta}$ *which maps into infinity by* $f(z)$. *As* $v$ *passes counterclockwise through* $e^{-i\alpha}$, *we suppose that* $\sqrt{Q(v)}$ *passes from negative to positive values. Then the function* $f^*(z)$ *belonging to any point* $\mathbf{a}^*$ *of* $M_n$ *in a sufficiently small neighborhood* $U(\mathbf{a}, \epsilon)$ *of* $\mathbf{a}$ *is given by the formula*

$$f^*(z) = f(z) - \epsilon\frac{f(z)^2}{2\pi i}\int_{|t|=1}\left(\frac{tf'(t)}{f(t)}\right)^2 \frac{\chi(t)}{\sqrt{Q(t)}\,[f(z) - f(t)]}\,\frac{dt}{t} + o(\epsilon),$$

*where the integration is counterclockwise around* $|t| = 1$ *beginning and ending at* $e^{-i\beta}$. *Here the function* $\sqrt{Q(t)}$ *is to be interpreted as above.*

---

* This paper was written while the authors were engaged on a project sponsored by the Office of Naval Research (Contract N6-ori, 154, Task III).

[1] Prawitz, H., "Ueber Mittelwerte analytischer Functionen," *Arkiv för Mat., Astron. och Fysik*, **20A**, 6, 1–12 (1927).

[2] Schaeffer, A. C., and Spencer, D. C., "The Coefficients of Schlicht Functions III," PROC. NAT. ACAD. SCI., **32**, 111–116 (1946).

---

# *ON ZERMELO'S AND VON NEUMANN'S AXIOMS FOR SET THEORY*

BY HAO WANG

SOCIETY OF FELLOWS, HARVARD UNIVERSITY

Communicated by J. von Neumann, January 22, 1949

If we assume the theory of truth-functions and the theory of quantifiers from the beginning[1] and disregard the axiom of choice, Zermelo's and von Neumann's systems of set theory may be described as follows.

The membership symbol $\epsilon$ connecting variables is taken as primitive in both systems so that any combination of the form $\ulcorner(\alpha \;\epsilon\; \beta)\urcorner$ where $\alpha$ and $\beta$ are any variables is an atomic formula and all further (well-formed) formulae are obtainable from atomic ones by binding quantifiers and applying truth-functional connectives.

In Zermelo's system abstracts (such as $\ulcorner\hat{x}\phi x\urcorner$, $\ulcorner\hat{y}\psi y\urcorner$, etc.), identity and inclusion may be introduced by the following definitions. Let $\ulcorner\phi x\urcorner$ be a formula in which $x$ is free but $y$ is not free, and $\ulcorner\phi y\urcorner$ be the result obtained from $\ulcorner\phi x\urcorner$ by substituting occurrences of $y$ for all free occurrences of $x$. Let $\zeta$ and $\eta$ each be either a variable which is not $x$ or an abstract.

ZM. $\ulcorner(x \;\epsilon\; \hat{y}\phi y)\urcorner$ or $\ulcorner\hat{x}(x \;\epsilon\; \phi x)\urcorner$ for $\ulcorner\phi x\urcorner$.

ZI. $\ulcorner(\zeta = \eta)\urcorner$ for $\ulcorner(x)(x \;\epsilon\; \zeta \equiv x \;\epsilon\; \eta)\urcorner$.

ZS. $\ulcorner(\zeta \subset \eta)\urcorner$ for $\ulcorner(x)(x \;\epsilon\; \zeta \supset x \;\epsilon\; \eta)\urcorner$.

The axioms of Zermelo's system as revised by Fraenkel and Skolem are the following:[2]

Z1. Axiom of extensionality. $(x = y \supset (x \in z \equiv y \in z))$.
Z2. Axiom of pairing. $(\exists w)(x)(x \in w \equiv (x = y \vee x = z))$.
Z3. Axiom of infinity. $(\exists w)((\exists u)(u \in w) \cdot (x)(x \in w \supset (\exists y)(y \in w \cdot x \subset y \cdot x \neq y)))$.
Z4. Axiom of summation. $(\exists w)(x)(x \in w \equiv (\exists y)(x \in y \cdot y \in z))$.
Z5. Axiom of the set of subsets. $(\exists w)(x)(x \in w \equiv x \subset z)$.
Z6. If $\ulcorner \phi xy \urcorner$ is a formula in which $x$ and $y$ are free but $z$, $w$ and $u$ are not free, then $\ulcorner (x)(y)(z)((\phi xy \cdot \phi xz) \supset y = z) \supset (\exists w)(x)(x \in w \equiv (\exists y)(y \in u \cdot \phi yx)) \urcorner$ is an axiom.

Von Neumann's system may be described by first introducing new variables $a$, $b$, $c$, $d$, $e$, etc., which take as values only classes capable of being members of the other classes (elements). These element variables can be introduced by contextual definitions such as:

NA. $\ulcorner (a)\phi a \urcorner$ for $\ulcorner (x)((\exists y)(x \in y) \supset \phi x) \urcorner$,
NE. $\ulcorner (\exists a)\psi a \urcorner$ for $\ulcorner (\exists x)((\exists y)(x \in y) \cdot \psi x) \urcorner$,

etc., where $\ulcorner \phi a \urcorner$ and $\ulcorner \psi a \urcorner$ are like $\ulcorner \phi x \urcorner$ and $\ulcorner \psi x \urcorner$ except for containing free occurrences of $a$ wherever $\ulcorner \phi x \urcorner$ and $\ulcorner \psi x \urcorner$ contain free occurrences of $x$. The definitions for abstracts, identity, and inclusion are:

NM. $\ulcorner (a \in \hat{b}\phi b) \urcorner$ or $\ulcorner (a \in \hat{a}\phi a) \urcorner$ for $\ulcorner \phi a \urcorner$.
NI. $\ulcorner (\zeta = \eta) \urcorner$ for $\ulcorner (a)(a \in \zeta \equiv a \in \eta) \urcorner$.
NS. $\ulcorner (\zeta \subset \eta) \urcorner$ for $\ulcorner (a)(a \in \zeta \supset a \in \eta) \urcorner$.

The indispensable axioms of von Neumann's system as revised by Bernays and Goedel are (with non-essential changes) the following:[3]

N1. $x = y \supset (x \in z \equiv y \in z)$.
N2. $(\exists d)(a)(a \in d \equiv (a = v \vee a = c))$.
N3. $(\exists d)((\exists e)(e \in d) \cdot (a)(a \in d \supset (\exists b)(b \in d \cdot a \subset b \cdot a \neq b)))$.
N4. $(\exists d)(a)(a \in d \equiv (\exists b)(a \in b \cdot b \in c))$.
N5. $(\exists d)(a)(a \in d \equiv a \subset c)$.
N6. $(e)(v)((a)(b)(c)((<ab> \in v \cdot <ac> \in v) \supset b = c) \supset (\exists d)(a)(a \in d \equiv (\exists b)(b \in e \cdot <ba> \in v)))$.
N7. If $\ulcorner \phi a \urcorner$ is a formula in which $y$ is not free and all bound variables are element variables, then $\ulcorner (\exists y)(a)(a \in y \equiv \phi a) \urcorner$ is an axiom.

If in the above system we replace N7 by a stronger principle due to Quine:[4] N7′. If $\ulcorner \phi a \urcorner$ is a formula in which $y$ is not free, then $\ulcorner (\exists y)(a)(a \in y \equiv \phi a) \urcorner$ is an axiom, we obtain a system NQ which appears to be somewhat stronger than von Neumann's system.[5] Unlike N7, N7′ is not yet known to be reducible to finitely many axioms. Nevertheless, this system

NQ seems to be an interesting extension of von Neumann's, because in it we may form classes without asking whether the defining predicate is a "constitutive expression,"[6] and the mere replacement of N7 by N7′ does not seem to affect the consistency of the system.[7]

Let Z1′–Z6′, ZM′, ZI′ and ZS′ be the result obtained from Z1–Z6, ZM, ZI and ZS by substituting element variables for all general variables. Then Z2′–Z5′ are just N2–N5, and Z1′, ZM′, ZI′ and ZS′ are special cases of N1, NM, NI and NS by dint of NA. Moreover, by N7, Z6′ is a consequence of N6. In fact, with the help of N7, we can deduce from N6 a principle N6′ somewhat stronger than Z6′: N6′. If $\ulcorner \phi ab \urcorner$ is a formula in which all bound variables are element variables and in which $a$, $b$ are free but $c$, $d$, $e$ are not free, then $\ulcorner (a)(b)(c)((\phi ab \cdot \phi ac) \supset b = c) \supset (\exists d)(a)(a \in d \equiv (\exists b)(b \in e \cdot \phi ba))\urcorner$ is an axiom. Consequently, to every class in Zermelo's theory there corresponds an element in von Neumann's theory, although there may be some element in von Neumann's system to which no class in Zermelo's system corresponds.

Obviously von Neumann's system is stronger than Zermelo's. More precisely, to every theorem on classes in Zermelo's system, there corresponds a theorem on elements in von Neumann's, because, by using definitions such as NA and NE, we can also prove that the result of substituting element variables for all general variables in any theorem of the theory of quantifiers is again a theorem in it. In particular, if there is a contradiction in Zermelo's system, there is also one in von Neumann's. In other words, if von Neumann's system is consistent, so also is Zermelo's.

On the other hand, it is not at all obvious that if Zermelo's system is consistent, so also is von Neumann's. Quine's system of *New Foundations*[8] is not known to be inconsistent, yet his system originally presented in ML, which is related to the former system in a similar manner as von Neumann's is related to Zermelo's, has been shown to be inconsistent.[9] Indeed, as von Neumann has observed,[10] it does seem that none of the known contradictions will arise in his system because, although the troublesome classes can be proved to exist, they are not elements in his system, and contradictions arise only when these classes are taken to be elements. But his remark certainly does not amount to a proof of the relative consistency of his system to Zermelo's. It remains to be demonstrated that his system is consistent if Zermelo's is.

In this paper I only propose to investigate how we can obtain von Neumann's system by a straightforward extension of Zermelo's. The result will bring out more clearly the point that if we regard Zermelo's theory as a functional calculus of the first order, von Neumann's system may be considered to be a functional calculus of the second order.

The extension is effected by first introducing into Zermelo's system a new kind of variables $X$, $Y$, $Z$, $U$, $W$, etc., and stipulating that the membership symbol $\epsilon$ when connecting a small variable to a large one be also taken as primitive. Accordingly, expressions such as $x \epsilon X$, $x \epsilon Y$, etc., are also atomic formulae, which, together with the other atomic formulae $x \epsilon y$, $x \epsilon z$, etc., may again generate new formulae. Large variables occurring before $\epsilon$ are introduced by definitions such as:

ZZM1. $(X \epsilon Y)$ for $(\exists z)(z = X \cdot z \epsilon Y)$,
ZZM2. $(X \epsilon y)$ for $(\exists z)(z = X \cdot z \epsilon y)$.

We further stipulate that the result of replacing all small variables in any theorem of the theory of quantifiers by these large variables is a theorem. The definitions ZM, ZI, ZS, and the principle Z6 are reconstrued as covering also the cases where $\ulcorner \phi x \urcorner$ and $\ulcorner \phi y \urcorner$ in ZM and $\ulcorner \phi xy \urcorner$ in Z6 contain large variables and one or both of $\zeta$ and $\eta$ are large variables or the new abstracts. These reinterpreted definitions and principles will be referred to as ZZM, ZZI, ZZS, ZZ6.

We add one more axiom and one rule of inference to Zermelo's system:[11]

ZZ1. $x = y \supset (x \epsilon X \equiv y \epsilon X)$. $x = Y \supset (x \epsilon z \equiv Y \epsilon z)$. $x = Y \supset (x \epsilon Z \equiv Y \epsilon Z)$. $X = Y \supset (X \epsilon z \equiv Y \epsilon z)$. $X = Y \supset (X \epsilon Z \equiv Y \epsilon Z)$.

ZZ7. If $x$ is free in $\phi$ and all the bound variables in $\phi$ are small variables $\psi'$ is like $\psi$ except for containing free occurrences of $\ulcorner \hat{x}\phi \urcorner$ wherever $\psi$ contains free occurrences of $X$, and $\psi$ is a theorem, then $\psi'$ is also a theorem.

In order to show that the result of thus enlarging Zermelo's system is "equivalent" to von Neumann's, we first correlate the large and small variables with the general and element variables, respectively. Then the last statement of ZZ1 answers to N1, Z2–Z5 answer to N2–N5, and by ZZM, ZZ6 and ZZ7, we can easily deduce ZZ6′ and ZZ7′ which are exact counterparts of N6 and N7. What remains to be shown is that we can prove biconditionals which answer to definitions such as NA and NE. Once we have proved such biconditionals, we can easily see that the two systems are equivalent, for (1) ZZM, ZZI, ZZS are then the exact counterparts of NM, NI, NS, (2) Z1 and the statements in ZZ1 become special cases of the last statement in ZZ1, (3) we can easily prove a principle answering to ZZ7 in von Neumann's system,[12] and (4) the biconditionals answering to ZZM1 and ZZM2 are easily provable by N1 and NE in von Neumann's system.

In order to prove the biconditionals corresponding to definitions such as NA and NE, we observe that the following statements are provable in the enlarged Zermelo system:[13]

T1. If $\ulcorner\phi X\urcorner$ is like $\ulcorner\phi x\urcorner$ except for containing free occurrences of $X$ in place of some free occurrences of $x$, then $\ulcorner x = X \supset (\phi x \equiv \phi X)\urcorner$ is a theorem.

T2. $(\exists Y)(x \in Y)$. By ZZ1, $x = x$. So, by ZZM, $x \in \hat{x}(x = x)$, and by ZZ7, $(\exists Y)(x \in Y)$.

T3. $(\exists Y)(X \in Y) \equiv (\exists x)(x = X)$. By ZZM1 and the theory of quantifiers, $(\exists Y)(X \in Y) \equiv (\exists x)(x = X \cdot (\exists Y)(x \in Y))$. Therefore, by T2, we obtain T3.

T4. $(\exists X)(x = X)$. By ZZM, $x = \hat{y}(y \in x)$. Therefore, by ZZ7, $(\exists X)(x = X)$.

Then we can prove a metatheorem corresponding to NA:

TNA. If $\ulcorner\psi X\urcorner$ is related to $\ulcorner\psi x\urcorner$ as $\ulcorner\phi x\urcorner$ is related to $\ulcorner\phi a\urcorner$ in NA, then $\ulcorner(x)\psi x \equiv (X)((\exists Y)(X \in Y) \supset \psi X)\urcorner$ is a theorem.

*Proof:* By T1, $x = X \supset (\psi x \equiv \psi X)$. But $(x)\psi x \supset \psi x$. Therefore, $x = X \supset ((x)\psi x \supset ((\exists Y)(X \in Y) \supset \psi X))$. So, $(\exists x)(x = X) \supset ((x)\psi x \supset ((\exists Y)(X \in Y) \supset \psi X))$. Hence, by T3,

$$(x)\psi x \supset (X)((\exists Y)(X \in Y) \supset \psi X). \tag{1}$$

On the other hand, by T1 and T2, $x = X \supset ((X)((\exists Y)(X \in Y) \supset \psi X) \supset \psi x)$. By T4,

$$(X)((\exists Y)(X \in Y) \supset \psi X) \supset (x)\psi x. \tag{2}$$

(1) and (2) taken together give us TNA.

The metatheorem for NE follows by the theory of quantifiers as a corollary.*

---

[1] For example, as are developed in the first three chapters of Hilbert-Ackermann, *Grundzuege der theoretischen Logik*, Berlin, 1938, or in the first two chapters of Quine, *Mathematical Logic*, New York, 1940. I shall assume Quine's version. Quine's book will be referred to as ML.

[2] See, for example, Ackermann, "Mengentheoretische Begruendung der Logik," *Math. Ann.*, 115, 1–22. Z6 takes the place of *Aussonderungsaxiom* and *Ersetzungsaxiom*, see footnote 9, page 9 of the cited paper. Compare also consequence 4 of Bernays's axioms IV, Va, Vb, page 3, *J. of Symbolic Logic*, 6.

[3] See Goedel, *The Consistency of the Continuum Hypothesis*, Princeton, 1940, pages 3–7. The principle N7 takes the place of Goedel's axioms B1–B8 plus an assertion that $(\exists y)(a)(a \in y \equiv a \in a)$; cf. M3 on page 14. $\langle xy\rangle$ is the ordered pair of $x$ and $y$ $(= \{\{x\}\{xy\}\})$.

[4] This is Quine's *202 in ML, page 162.

[5] The difference between N7 and N7′ in their deductive power is not trivial. For example, we can obtain the theory of natural numbers with N1, N2 and N7′ plus the assertion that the empty class is (represented by) an element (cf. *J. of Symbolic Logic*, 13, p. 129), but the same construction cannot be carried through if we use N7 instead of N7′, because the defining formula for the ancestral relation (ML, p. 216) contains a bound variable which is not an element variable.

[6] See, for example, top of page 9, *J. of Symbolic Logic*, **6**.

[7] Compare middle of page 158, *Ibid.*, **7**.

[8] *Am. Math. Mo.*, **44**, p. 70.

[9] *J. of Symbolic Logic*, **7**, p. 1. The inconsistent system contains N7′ in place of N7. However, whether the replacement of N7′ by N7 would affect the derivation of the Burali-Forti contradiction is not immediately obvious.

[10] Middle of page 228, *J. r. angew. Math.*, **160**.

[11] If we want to obtain N7′ instead of N7, just delete the clause "and all the bound variables in $\phi$ are small variables" in *ZZ7*.

[12] *Compare the proof of *231, ML, page 171.*

[13] *T1 can be proved by ZZI and ZZ1 in the usual manner.* Cf., e.g., *ML*, page 170, *223.

* I am indebted to Dr. I. L. Novak for suggestion in connection with the introduction of large variables. I wish also to thank Dr. Novak and Dr. H. Hiz for reading the manuscript of the paper and suggesting improvements in the manner of presentation. I am grateful to Professor Quine for corrections.

---

# *THE CULTURE OF PARAMECIUM AURELIA IN THE ABSENCE OF OTHER LIVING ORGANISMS*

BY W. J. VAN WAGTENDONK AND PATRICIA L. HACKETT

DEPARTMENT OF ZOOLOGY, INDIANA UNIVERSITY, BLOOMINGTON, INDIANA

Communicated by T. M. Sonneborn, January 22, 1949

The growth requirements of ciliated protozoa are very complex, and little accurate work on this has so far been possible since only a few species have been cultivated in the absence of other living organisms. Sterile cultures of *Tetrahymena geleii*,[1] *Colpoda duodenaria*[2, 3] and *Paramecium multimicronucleatum*[4] have been successfully established. *T. geleii* is thus far the only ciliated protozoan which has been cultured in a synthetic medium in the absence of natural proteins or peptones.[5] In addition to an as yet unidentified "plasmoptyzate factor," presumably of protein nature,[3] *C. duodenaria* requires thiamin, pyridoxin and nicotinic acid in the medium.[6, 7] The growth requirements of *P. micronucleatum* have not yet been established. It has been reported that this organism requires a heat-labile and a heat-stabile factor, both present in yeast press juice and in a bacterial plasmoptyzate.[8]

*P. aurelia* is routinely grown in a lettuce infusion inoculated with *Aerobacter aerogenes*.[9] While this procedure is satisfactory for general purposes, a standardization of this culture medium would be desirable. This requires the establishment of a culture of *P. aurelia* in a medium devoid of other living organisms, which would sustain the growth of this ciliate at a satisfactory fission rate. After such a medium is developed, a study can be made of the specific growth requirements of this organism.

*Sterilization of P. aurelia.*—Two different procedures were used to free *P. aurelia* from other living organisms: (*a*) electrical transport of the ciliates against a continuous flow of a sterile agar solution, and (*b*) exposure of *P. aurelia* to penicillin solutions.

(*a*) *Sterilization by Electromigration:* The electromigration tube is shown in figure 1. The tube and the Erlenmeyer containing the wash fluid (NaCl 0.0005 g., $KH_2PO_4$ 0.008 g., $MgCl_2 . 7\ H_2O$ 0.0035 g., $FeCl_3 . 6\ H_2O$ 0.006 g., $CaCO_3$ 0.0005 g., agar 1.0 g., lettuce infusion 10 ml., distilled water to make 1 liter) were sterilized separately by autoclaving. Immediately before use the two parts were connected asceptically, and the tube was filled by opening stopcocks *A*, *B* and *C*. The flow of the liquid through the tube was adjusted to a rate of 0.6 ml. per minute. The electrodes were connected to a 9-volt battery, thereby producing a current of 0.1 to 0.2 ma. A suspension of *P. aurelia* was injected into side arm *D*. The average migration time of the ciliates from *D* to *E* was about 27 minutes. When the first *P. aurelia* appeared at the side tube *E*, they were removed by means of a sterile syringe and placed in sterile culture fluids of different composition. At the same time samples were distributed to various culture media in order to check for sterility. Subsequent samples of *P. aurelia* can be removed from the tube at point *E* until all of the wash fluid is exhausted. From 30 to 50 ciliates can be recovered from each sample. Under the described conditions about 70 per cent of the samples are sterile.

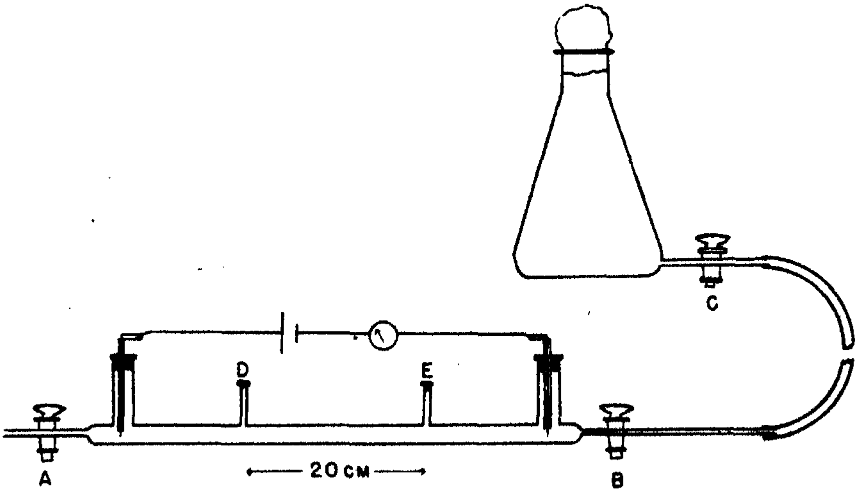


FIGURE 1

Electromigration tube.

(*b*) *Sterilization with Penicillin:* Seaman[10] was the first to use penicillin as an agent for the sterilization of protozoans. A wild culture of *C.*

*campylum* was sterilized by treatment with 5000 units of penicillin per ml. of a 5 per cent proteose peptone broth for 12 hours. In preliminary experiments it was found that *P. aurelia* cannot tolerate concentrations of penicillin higher than 3200 units per ml.

Approximately 20 *P. aurelia* were washed through nine 1-ml. volumes of sterile lettuce infusion, and the small number of organisms which were recovered from the washing process were grown in lettuce infusion inoculated with *A. aerogenes* until the population had reached a concentration of 2000 animals per ml. These paramecia, contaminated by the presence of only a single organism, were then transferred to a solution of 1 per cent yeast autolyzate (Basamin-Busch, kindly furnished by the Anhaeuser Busch Company) saturated with $CaCO_3$, and containing 3200 units of penicillin G-sodium salt per ml. (obtained through the courtesy of Dr. J. M. McGuire of Eli Lilly and Company). This solution had been sterilized by filtration through a sintered glass bacterial filter. After 48 hours in this medium, the organisms were washed through 5 2-ml. volumes of sterile lettuce infusion. The *P. aurelia* from the last wash were then transferred to various media. At the same time samples were distributed to various culture media in order to test for sterility.

(*c*) *Sterility Tests:* *P. aurelia*, treated in either of the two ways described under (*a*) or (*b*), were tested for the presence of contaminating organisms. Aliquots of 0.5 ml. of culture medium to which the organisms had been transferred after the sterilization procedure (taken immediately after the transfer), and aliquots of 0.5 ml. of the culture medium after *P. aurelia* had been in this medium for 5 days were inoculated in the following media: (1) yeast beef agar (Difco), (2) 0.5 per cent yeast extract (Difco), (3) 0.5 per cent yeast extract to which 0.5 per cent sucrose had been added, (4) 0.5 per cent beef extract (Difco), (5) nutrient agar (Difco) to which 0.5 per cent sucrose had been added. The tests were incubated at three different temperatures, 20°, 27°, 37°, under both aerobic and anaerobic conditions. If no contamination occurred, as observed by the absence of cloudiness in the liquid media or absence of colonies on the solid media, in three weeks, the originally isolated *P. aurelia* were considered to be free of any contamination. The organisms were independently tested for sterility by Dr. L. S. McClung of the Department of Bacteriology of Indiana University who confirmed our observations.

*Culture in Sterile Media.*—Of the various media tried, only the following were able to sustain growth of *P. aurelia:*

(*a*) Equal volumes of 0.5 per cent yeast autolyzate (Basamin-Busch) sterilized by filtration through a sintered glass bacterial filter, and an autoclaved lettuce infusion (20 minutes, 15 pounds), which had been inoculated with *A. aerogenes* 24 hours previous to autoclaving.

(*b*) Equal volumes of 0.5 per cent yeast autolyzate, sterilized by autoclaving (20 minutes, 15 pounds) and an autoclaved lettuce infusion which had been inoculated with *A. aerogenes* 24 hours previously.

The organisms would multiply in this medium at an average fission rate of 1.7 fissions per day (table 1). The culture has been carried through more than 30 successive transfers. At this time no contamination could be detected by any of the methods mentioned under (*c*). Thus far it has not been possible to make any substitutions for either the particular brand of yeast autolyzate or for the autoclaved bacterial suspension. Various other brands of dried yeast preparations and preparations of fresh yeast were tried but none of these could support the growth of *P. aurelia*. Addition of a bacterial autolyzate in 1 *M* glycine[11] did not allow *P. aurelia* to survive for more than 10 days. No fission took place in this interval. Ultrasonic preparations of an Aerobacter suspension, followed by sterilization by filtration failed to yield preparations which, when added to the yeast autolyzate, would sustain the growth of the ciliates. When either the lettuce infusion or the autoclaved bacterial culture was omitted, the medium became incapable of supporting the growth of the ciliates; similar results were obtained when the three-component medium (described under (*a*) or (*b*) was filtered through a sintered glass bacterial filter. Rapid death resulted when the autoclaved lettuce–Aerobacter infusion was the only constituent of the medium.

TABLE 1

GROWTH RATE OF *Paramecium aurelia* IN A STERILE MEDIUM

| NO. OF 48-HR. TRANSFERS | INITIAL NO. OF *P. aurelia* | NO. OF *P. aurelia* AFTER 48 HRS. | FISSIONS PER DAY |
|---|---|---|---|
| 1 | 81 | 563 | 1.4 |
| 2 | 56 | 371 | 1.4 |
| 3 | 37 | 272 | 1.4 |
| 4 | 85 | 680 | 1.5 |
| 5 | 56 | 681 | 1.8 |
| 10 | 22 | 350 | 2.0 |
| 15 | 51 | 690 | 1.8 |
| 20 | 41 | 524 | 1.8 |
| 25 | 74 | 1170 | 2.0 |
| 30 | 15 | 216 | 1.9 |

*Discussion.*—In view of the accumulated knowledge of the nutrition of exacting species of protozoans,[12, 13] it is surprising that *P. aurelia* is capable of growing and multiplying in a medium containing only heat-stabile components. However, the fact that the sterile medium permits only a maximal growth of 1.7 fissions per day, as compared with 4 to 6 fissions per day in a lettuce–Aerobacter infusion, suggests that heat-labile factors may be necessary for optimal growth. The complex nutritional requirements of *P. aurelia* are evident from the fact that a three-component medium is required for sterile growth. It should now be possible to fractionate the different components into their constituents and thus arrive at a synthetic medium for this organism. All three components of the medium, yeast extract, lettuce infusion and the autoclaved bacterium *A. aerogenes*, contribute essential factors for the nutrition of *P. aurelia*. The observation that this organism will not grow in a medium composed of the three essential components, but filtered through a bacterial filter seems to indicate that some particles of large size are also needed for its growth. However, it might be possible that some essential nutrilite is so firmly bound to the bacterial cell constituents that it had not been liberated in solution by the methods used.

Contribution No. 398 from the Department of Zoology, Indiana University. The work was supported by grants from the Jane Coffin Childs Memorial Fund for Medical Research, the National Institute of Health of the United States Public Health Service and the Rockefeller Foundation (grant for research on Paramecium genetics).

[1] Lwoff, A., *C. R. Acad. Sci. (Paris)*, **176**, 928 (1923).

[2] Taylor, C. V., and van Wagtendonk, W. J., *J. Cellular Comp. Physiol.*, **17**, 349 (1941).

[3] van Wagtendonk, W. J., and Taylor, C. V., *Ibid.*, **19**, 95 (1942).

[4] Johnson, W. H., and Baker, E. G. S., *Science*, **95**, 333 (1942).

[5] Kidder, G. W., *Ann. N. Y. Acad. Sci.*, **69**, 99 (1947).

[6] Tatum, E. L., Garnjobst, L., and Taylor, C. V., *J. Cellular Comp. Physiol.*, **20**, 199 (1942).

[7] Garnjobst, L., Tatum, E. L., and Taylor, C. V., *Ibid.*, **21**, 211 (1943).

[8] Johnson, W. H., and Tatum, E. L., *Arch. Biochem.*, **8**, 163 (1945).

[9] Sonneborn, T. M., and Dippell, R. V., *Physiol. Zoöl.*, **19**, 1 (1946).

[10] Seaman, G. R., *Science*, **106**, 327 (1947).

[11] Maculla, E. S., and Cowles, P. B., *Ibid.*, **107**, 376 (1948).

[12] Hall, R. P., *Protozoa in Medical Research*, University of Columbia Press, 1941, pp. 474–516.

[13] Hall, R. P., *Vitamins and Hormones*, Vol. 1, 1943, p. 249.

PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES

Volume 35 April 15, 1949 Number 4

# SOME PROBLEMS OF THE EVOLUTION OF STARS

By Otto Struve

Yerkes Observatory, Williams Bay, Wisconsin, University of Chicago

To be read before the Academy, April, 1949

During the past few years several stimulating articles have appeared outlining a new theory of the evolution of stars.[1] The purpose of this note is to present certain considerations regarding the evolution of close binary stars and to discuss them in the light of the more general theories, especially that of v. Weizsaecker. The earlier theories of double-star formation are so well known that they need not be reviewed in this place. The most convincing arguments in favor of a fission process have probably been advanced by James Jeans, who in many respects continued the earlier work of George Darwin. However, serious doubts regarding the possibility of fission were expressed many years ago by F. R. Moulton, and the most potent considerations against the classical fission theory were published about fourteen years ago by G. P. Kuiper.[2] The problems of cosmogony are so involved and difficult that it is important to record the considerations of all those who have worked on this subject. Most investigators are primarily influenced by the results obtained within the scope of their own investigations. This leads me to call attention to certain observational results which are not very widely known but which have an important bearing upon all cosmogonical speculations. I am referring here particularly to the results obtained from studies of stellar rotation.

V. Weizsaecker's recent theory is essentially based upon his earlier work on the effect of turbulence in diffuse stellar objects which, in keeping with the terminology of Spitzer and Whipple, may be designated as protostars. These protostars are believed to have originated through the condensation of interstellar clouds of gas and smoke, very much as has been recently suggested by Spitzer[3] and Whipple.[4] Turbulent motions in the original medium would have produced large angular momenta in most of these protostars, and this presumably accounts for the high frequency of rapidly rotating stars of the earlier spectral types which, according to this theory, are relatively young. As the result of turbulent motions within the protostar, whose rotation is not that of a rigid body, angular momentum is transported from the interior regions to the outside, and finally the mass breaks into two, the nucleus retaining only a small amount of angular momentum while the outer fringe of the diffuse, nebulous mass is dispersed into space, carrying away in the process the excess angular momentum. In this manner v. Weizsaecker explains the gradual loss of mass in a hot main-sequence star and the absence of large rotational velocities in the stars of later spectral types.

However, it is now known from the work of H. Shapley[5] that although the overwhelming majority of the cooler stars of the main sequence in the Hertzsprung-Russell diagram possess only a very small amount of angular momentum there are in existence a large number of binaries of late spectral type which possess rapid rotations and large amounts of orbital angular momentum. Shapley has shown that these so-called W Ursae Majoris binaries are more numerous in space than all the other variable stars now recognized. Although he does not give a quantitative estimate of the fraction of solar type stars which are binaries of this kind it is clear that they are sufficiently numerous to require consideration when we examine the evolution of the stars in the main sequence.

A W Ursae Majoris binary normally has a spectral type close to that of the sun. Its surface temperature is therefore in the vicinity of 6000°. Each system consists of two components having masses of the order of 1 solar mass and $^1/_2$ solar mass, respectively. The diameters of these stars are somewhat smaller than those of the sun. They revolve almost in contact and they are surrounded by a common gaseous envelope which is unsymmetrical in thickness and in other physical properties.[6] The average orbital velocities of the approximate geometrical centers of the two components with respect to the center of gravity of the whole system are 110 km./sec. and 220 km./sec. Since the system rotates approximately as a rigid body the relative velocity of the points of greatest separation on the two stellar ellipsoids is at least 400 km./sec. Hence, if in the course evolution such a star should ever combine to form a single object, its rotational velocity would be considerably more than 200 km./sec. at the equator. Such a velocity would be easy to observe. Yet among the single stars of appropriate spectral type there are no objects having anything like this kind of rotation.

It might be asked what kind of star would result from the coalescence of the two components. There are all reasons to think that the process would be a fairly gradual one so that any disturbance in the constitution of the object would not for very long remove that star from the main sequence. The existence of a common envelope makes it, even now, difficult to decide whether they are single stars with two condensations, or real binaries. At any rate, if the two stars should combine without losing mass they would form a single object having a mass of about 1.5 times the mass of the sun. If such a star remained on the main sequence, its spectral type would be late F or early G and observations show that there are no single stars within these spectral subdivisions which have the required amount of angular momentum. If the star should be disturbed to such an extent as to depart from the main sequence it would have to fall somewhere along the appropriate line for mass 1.5 $\odot$ in the H-R diagram. Otherwise it would be necessary to suppose that it would fail to obey the mass luminosity relation and this would place it in a portion of the H-R diagram where there are so few stars per unit volume of space that we can safely exclude the possibility. As a matter of fact this argument strongly suggests that the very numerous W Ursae Majoris type binaries do not have descendant single stars whose mass is 1.5 $\odot$. Similarly we can also show that the reverse process is not possible because none of the single stars of mass 1.5 $\odot$ has enough angular momentum to produce a binary of the W Ursae Majoris type.

TABLE 1

Systems of the Types of W Ursae Majoris

| | | | P, | KM./SEC. | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| NO.* | STAR | SP. | DAYS | $\gamma$ | $K_1$ | $K_2$ | $m_1$ | $m_2$ | $\alpha$ | INT. | $J_1/J_2$ |
| 1 | W UMa | F8 | 0.334 | − 5 | 134 | 188 | 0.9$\odot$ | 0.6. | 1.4 | Var. | <1 |
| 2 | TX Cnc | F8 | 0.383 | +36 | 112 | 217 | 1.5 | 0.8 | 1.9 | Var. | <1 |
| 3 | 44i Boo B | G2 | 0.268 | + 3 | 116 | 232 | 1.0 | 0.5 | 2.0 | ? | <1 |
| 4 | ER Ori | G2 | 0.423 | +35 | 95 | 155 | 0.6 | 0.4 | 1.6 | ? | >1 |
| 5 | YY Eri | G5 | 0.321 | −20 | 130 | 200 | 1.0 | 0.7 | 1.5 | Var. | <1 |
| 6 | VW Cep | KO | 0.278 | −35 | 75 | 230 | 1.1 | 0.35 | 3.1 | Var. | <1 |
| 7 | AH Vir | KO | 0.408 | +10 | 105 | 250 | 1.5 | 0.7 | 2.4 | Var. | <1 |
| 8 | RZ Cam | KO | 0.339 | −12 | 130 | 270 | 1.6 | 0.8 | 2.1 | Var. | <1 |
| 9 | V 502 Oph | G2 + F9 | 0.453 | −37 | 95 | 235 | 1.2 | 0.5 | 2.5 | ? | <1 |
| 10 | SW Lac | G3 | 0.321 | −22 | 172 | 202 | 0.5 | 0.4 | 1.2 | Var. | <1 |
| | mean | | | | 110 | 220 | 1.00 | 0.5 | 2 | | |

* References: 1. Joy, *Astrophys. J.*, **49**, 189 (1919); Struve and Iwanowska (unpublished); 2. Popper, *Ibid.*, **108**, 490 (1948); 3. Popper, *Ibid.*, **97**, 407 (1943); 4. Struve, *Pubs. Astron. Soc. Pacific*, **56**, 34 (1944); 5. Struve, *Astrophys. J.*, **106**, 92 (1947); 6. Popper, *Ibid.*, **108**, 490 (1948); 7. Chang, *Ibid.*, **107**, 96 (1948); 8. Struve and Gratton, *Ibid.*, **108**, 502 (1948); 9. *Ibid.*; 10. Struve, *Ibid.* (in press).

I believe that this calls for a modification of v. Weizsaecker's theory. Otherwise the W Ursae Majoris type binaries would remain unexplained and they are altogether too numerous to be ignored.

Spectroscopic observations are now available for ten of these objects. The pertinent data are summarized in table 1.

The third column of the table gives the spectral type which is the same for both components of each system except in the case of V 502 Ophiuchi. In that star the more massive component is of spectral type G2 and the less massive of type F9. The fourth column gives the period in days. It ranges from 0.268 day to 0.453 day. The fifth column gives the radial velocity in km./sec. of the center of gravity of each system. The ten individual values clearly reflect the motion of the solar system through the local group of stars and in addition shows that the individual peculiar motions of these binaries are distributed approximately as they are among single stars of similar spectral types and luminosities. The sixth and seventh columns give the orbital velocities of the two components in km./sec. with respect to the center of gravity. Columns 8 and 9 list the masses of the two components in units of the mass of the sun. The quantity $\alpha$ is the mass ratio. The next to the last column shows whether the spectral lines of the fainter component undergo variations in intensity and the last column shows whether the ratio of the surface brightnesses of the more massive component to the less massive component is smaller than 1 or greater than 1. Attention should be called to the fact that the spectroscopic observations give us only the mass functions of the two components $m_1 \sin^3 i$ and $m_2 \sin^3 i$. The inclinations are either known from the photometric observations or they can be roughly estimated, but there is some uncertainty attached to these values of $i$. Hence the tabular masses of the two stars are not always known with great precision, but they are known sufficiently well to give us a general idea of the properties of these systems. On the contrary, the mass ratio $\alpha$ is known with great precision since it is equal to the ratio $K_2/K_1$.

Of special interest is the remarkable tendency of these systems to show spectral lines of variable intensity. As a general rule the approaching component is relatively stronger than the receding component. There are indications that the amount of variation is variable in time, but as yet no periodicities have been found. It is very probable that these changes are related to peculiar changes in the light curves of these variables which have recently been discussed by Eggen.[7] The two maxima of the light curves, which correspond to the times of greatest elongation of the two components, are not of the same height; and the amplitude of this variation, which can be represented by a term in sin $\theta$, where $\theta$ is the longitude in the orbit counted from primary minimum, is variable between zero and a small negative value. The exact connection between the photometric and spectrographic results is not yet known but, generally speaking, the system is less luminous when the less massive component is approaching and has its greatest absorption intensity. This phenomenon cannot be adequately explained as the result of ordinary differences in the temperatures of the advancing and receding hemispheres of the two stars and it must be somehow caused by the stream of gas which forms the common envelope.

Exactly the same spectroscopic phenomenon was found many years ago in spectroscopic binaries of very early type. It is most conspicuous in J. S. Plaskett's star of large mass, HD 47129, and it is also present in AO Cass, another very massive star of spectral type O. There can be no doubt that the phenomenon is the same and that in these two otherwise very different groups of binary systems we observe the effect of gaseous masses whose optical thickness is greatest over the advancing hemisphere of the less massive component of each system and perhaps also over the advancing hemisphere of the more massive component. The origin of these gaseous masses can only be guessed by analogy with several other binaries, especially UX Mon. I am inclined to believe that they may be caused by vast prominences which not only produce the variation in the line intensities but also give rise to the gaseous streams which influence the systems and produce various spectroscopic effects, such as double emission lines in numerous binaries of early spectral type and peculiar absorption features in other systems, which distort the velocity curves of these objects.

Some time ago I made a crude estimate of the amount of the material contained in such a ring-like gaseous formation. The result was $10^{-8}$ ⊙. No doubt, these streams of gas are secularly unstable. In the earlier and hotter systems radiation pressure must rapidly drive them out into space. We actually observe this phenomenon in $\beta$ Lyrae. Even if we assume that the rate of dissipation of the gaseous envelope results in its complete renewal once per year the original mass of the star would be exhausted in something like $10^{+8}$ years. This rate of dissipation would require an average velocity of expansion of only about $5 \times 10^{-2}$ km./sec. A rate of expansion 200 times faster, or 10 km./sec., would still be consistent with the absence of any conspicuous velocities of expansion in most systems (but in Plaskett's star the observed rate of expansion is about 100 km./sec. and in $\beta$ Lyrae it is of the order of 50 km./sec.). It is thus entirely possible that the mass of an early-type system would be nearly exhausted in an interval of $10^6$ years.

It is remarkable that v. Weizsaecker's estimate based upon his ideas of turbulence leads to precisely the same interval of time for the dissipation of mass and rotational momentum in a protostar of early spectral type and of an equatorial velocity of rotation of the order of 100 km./sec.

This leads me to suggest that perhaps the W Ursae Majoris systems have been formed out of the binaries of early type, such as $\beta$ Lyrae, U Cephei, or Plaskett's massive star. I recognize that this hypothesis has certain weaknesses. For example, there is the difference in the galactic concentration of the early-type binaries and the W Ursae Majoris systems and also the prevalence of larger individual space motions among the latter group. However, this difficulty is common to the present hypothesis as well as to v. Weizsaecker's original theory. It requires the assumption that in the course of $10^6$ or $10^8$ years the distribution of the stars in the galaxy and the approach to equipartition of the translational energies has been sufficiently accomplished and this would presumably require a much greater density of stars in the past than is observed at the present time. There is also the difference in the space frequencies of the early-type binaries and the binaries of the solar type. I am disregarding here those binaries which contain super-giant components. They are so rare in space as not to influence the arguments appreciably. The relative frequencies of close binary systems among the stars of types O, B and A with respect to the single stars of the same types is probably very nearly the same as for the W Ursae Majoris systems and the single stars of types: late F, G and K; but per unit volume of space the early-type stars are at least 50 times less frequent than are the late-type stars. Hence, the same ratio must apply also to the two groups of binaries. If the early-type binaries gradually lose mass and become W Ursae Majoris binaries it is necessary to assume that the length of time they remain in the latter stage greatly exceeds that which they require in their early-type stage. This argument again is not limited to my suggestion but is common also to v. Weizsaecker's theory, and it may be related to the relatively great youth which the Bethe theory of stellar energies attributes to all stars of very early type.

One further argument which suggests that the early-type binaries represent the youngest stage of double-star formation may perhaps be seen in the fact that those systems which, like U Cephei, consist of an early-type component of small size and a late-type component of large size show large departures from the mass luminosity relation for the fainter components: these sub-giants are much too luminous for their masses. It almost seems as though the less massive component in a recently formed binary system has been seriously disturbed from equilibrium and has wandered off the main sequence in the direction of higher luminosities. We cannot trace their course in the conventional H-R diagram because these stars do not obey the mass luminosity relation, but I think we can reasonably conclude that they are anomalous, even though the more massive early-type components are still on the main sequence. What happens when a star divides into two approximately equal components is not certain, because the empirical mass-luminosity relation at its upper end is based upon these systems. But it is significant that such stars as HD 47129 and AO Cass show spectroscopic features of very high luminosity. Unsöld and especially Weizsaecker have called attention to the possibility that some catastrophic phenomenon, such as a super-nova, results in the process, but perhaps this is carrying us too far since it is not necessary for our purpose to discuss whether some modernized version of the fission theory involving the interplay of nuclear reactions in the stellar interiors might meet the objections which Kuiper had made to the classical fission theory of Jeans.

One other point remains to be considered. What happens to the W Ursae Majoris binary if it continues dissipating mass and angular momentum through the medium of the common envelope, which in many respects would play the rôle of v. Weizsaecker's nebulosity? Perhaps what would result is precisely what v. Weizsaecker has predicted: a system of planets and a rapidly dissipating nebulosity which carries off most of the rotational momentum. If this is possible then it is perhaps not too far-fetched to think of the W Ursae Majoris systems as being the parent bodies of what we now observe as single solar-type stars without much angular momentum.

Despite its obviously bizarre appearance, this hypothesis has the attractive feature of giving us some observational evidence of precisely the kind of nebulosity which v. Weizsaecker requires. Otherwise, within the domain of main-sequence stars later than about type A, shell-like formations are not known, at least not among the single stars. Yet one would expect to observe various kinds of dilution effects in a mass of gas having the properties of the v. Weizsaecker nebulosity.

[1] V. Weizsaecker, C. F., *Z. Astrophys.*, **24**, 181 (1947); Unsöld, A., *Ibid.*, **24**, 278 (1948); Ambartzumian, A., "The Evolution of Stars and Astrophysics" (in Russian), *Acad. Sci. Armenian S.S.R.*, 1947.

[2] Kuiper, G. P., *Pubs. Astron. Soc. Pacific*, **47**, 15, 121 (1935).

[3] Spitzer, L., Jr., "Centennial Symposia Volume," *Harv. Obs. Monographs*, No. 7, page 87 (1948); "Physics Today," 1, No. 5, 6 (September, 1948).

[4] Whipple, F., "Centennial Symposia Volume," *Harv. Obs. Monographs*, No. 7, page 109 (1948).

[5] Shapley, H., "Centennial Symposia Volume," *Ibid.*, page 249.

[6] Kuiper, G. P., *Astrophys. J.*, **108**, 541 (1948); Struve, O., *Ann. d'Astrophysique*, **11**, 117 (1948).

[7] Eggen, O. J., *Astrophys. J.*, **108**, 15 (1948).

---

# *THE ACTION OF ALLELIC FORMS OF THE GENE A IN MAIZE II. THE RELATION OF CROSSING OVER TO MUTATION OF $A^b$* *

BY JOHN R. LAUGHNAN

DEPARTMENT OF BOTANY, UNIVERSITY OF ILLINOIS

Communicated by M. M. Rhoades, February 15, 1949

Inherited changes other than those which may be regarded as true mutations of the gene occur frequently in plant and animal material. Among such changes in *Drosophila* some are associated with crossing over between intrachromosomal duplications.[1-3] These latter may be distinguished from chromosomal aberrations of a more gross nature since the changes are localized within small regions and the duplicated segments involved carry either the same allele or factors closely related in their effects. It has been suggested that this highly localized duplication of chromatin followed by mutation and selection may constitute an important source of new genes. Further, depending on its frequency of occurrence, this phenomenon has a direct bearing on the interpretations of studies inquiring into the manner of action of genes. These studies assume, as the simplest situation, that each member of an allelic series is a single unit of action. If this assumption is maintained in those cases in which the effects of alleles are non-linear it becomes necessary to consider the action of this single unit as correspondingly more complex from a physiological viewpoint. However, if it can be demonstrated that the alleles in question are represented on the chromatin thread by separable entities it is possible to view their action according to schemes which are not so complex from the standpoint of action of the individual unit.

Evidence is presented here for the compound nature of the $A^b$ gene, a member of the relatively extensive $A$ series of alleles in maize. As noted in the original report,[4] in its determination of brown pericarp this allele is dominant to the wild type $A$, which is associated with red pigmentation. Since $A$ itself is dominant to recessive $a$ (brown pericarp) the dominant effect of $A^b$ is not expected on the basis of any simple scheme representing the action of these alleles. From a study[5] of the dosage and dominance effects of certain mutants derived from $A^b$ and designated $A^d$ it is known that these alleles retain the distinctly antimorphic effects of the parent allele; their action is unexplained on a simple basis.

$A^b$ mutates spontaneously with high frequency to alleles of lower level. Measured in the female germ line, Stadler[6] found a rate of 0.0005 (19: 36,661) for mutation to the $A^d$ allele. Among 55,765 tested female gametes, no mutations of $A^b$ to $a$ were found. However, mutations to $a$ occur indirectly. $A^d$, itself a mutant derivative of $A^b$, has produced 4 mutations to $a$ (4: ca. 50,000).[7] It appears from this that either the mutation of $A^b$ is highly directed or this gene may be compound. In the event of the latter the production of the mutant $a$ by two successive changes, but never in a single step, is explained if it is assumed that two components of $A^b$ are independent in their mutation.

To test the possibility that $A^b$ is represented by closely linked determinants experiments were undertaken to determine whether there is a relation between mutation of this allele and crossing over in or near the $A$ locus. For such an experiment it is desirable to study mutation in $A^b$ $A^b$ plants heterozygous for marker genes which are closely linked and located on either side of $A^b$. If a choice of marker genes can be made such that only a small segment of the chromosome lies between them, multiple exchanges in this region may be eliminated or greatly reduced. Under these conditions the frequency of strands which experience multiple exchanges in the marked segment but which appear as non-crossover products is negligible. Where the distance between marker genes is greater the efficiency of the experiment is reduced in proportion to the frequency of double crossover strands which appear as non-crossovers. Unfortunately, until very recently there were no known mutant genes showing close linkage with *A*. At the time the present experiments were begun the best choice of markers was that of $lg_2$ and *et*, which lie, respectively, 33 units to the left and 13 units to the right of *A* in the long arm of the third chromosome. Since the recombination value between *lg* and *et* is high (ca. 42%), it was considered that even in the event that mutations of $A^b$ were always associated with crossing over many of the mutants would appear as non-crossovers owing to the high frequency of double exchanges between *lg* and *et*. This disadvantage may be lessened if plants of the constitution $A^b\,a$ are employed instead of the homozygotes. This substitution allows the use of *a* as a third marker gene in the experiment; instead of 42 units of recombination the values corresponding to the *lg*-*a* and *a*-*et* intervals may now be employed. Since mutants of the $A^d$ type have never been obtained from *a*, it is certain, after suitable testing, that those which occur in these experiments have originated from $A^b$. Moreover the $A^d$ mutant seeds which are pale in appearance are distinguishable from both the deep purple seeds associated with $A^b$ and the colorless phenotype associated with *a*.

The data reported here are derived mainly from one type of cross: $A^b/lg\ a\ et \times lg\ a\ et$. Plants having these constitutions were grown at Princeton in the summers of 1947 and 1948. In order to avoid the occurrence of pale seeds due to contaminating pollen grains the plants were started and reached the flowering stage earlier than any other families in the field; moreover the mass pollination technique was discarded in favor of individual hand pollinations. The $A^d$ mutants from the 1947 experiments have had two generations of testing since their occurrence. These will be considered separately from those obtained in 1948 since the latter have not been as thoroughly tested.

Approximately half of the seeds produced by this cross have deep purple aleurones ($A^b\ a\ a$) and half are colorless ($a\ a\ a$). Ears of two families which produced a total of 27,936 purple seeds in 1947 were examined for the occurrence of pale seeds. Fifty-one pale seeds (suspected $A^d$ mutants) were obtained for testing. Of these, 36 had etched endosperms (*et et et*) and 15 were normal (*Et et et*). Not all were expected to be true mutants since these included many seeds which deviated only slightly in phenotype from the purple class. Unfortunately only twenty of the 51 cases survived beyond the seedling stage. Thirteen of these were indistinguishable at maturity from sib $A^b\ a$ plants grown as controls. That these plants carried an unmutated $A^b$ gene is established from their crosses with sib $A^b\ a$ plants. If the plants in question had carried a mutant $A^d$ gene ($A^d\ a$) three types of seeds would have resulted from these crosses: purple, pale and colorless. Only purple and colorless seeds were produced.

The remaining seven plants of the twenty which grew to maturity had a reddish brown plant color in distinct contrast to the purple phenotype of sib $A^b\ a$ plants. That these plants carried a mutated allele of $A^b$ producing an intermediate phenotype is confirmed by their selfed and outcrossed progenies. These data are summarized in table 1.

Six of the seven plants yielded progenies following self pollination. As indicated in the table only pale and colorless seeds were produced on these ears. Crosses of each of the pale plants with $A^b\ a$ individuals gave purple, pale and colorless seeds. From this it is apparent that all seven of the pale plants were heterozygous for a mutant gene associated with intermediate plant and aleurone phenotype and originating in the female germ line of the $A^b\ a$ plants of the original cross.

TABLE 1

SUMMARY OF DATA ON ALEURONE COLOR FROM SELFED AND CROSSED PROGENIES OF THE SEVEN PALE PLANTS ORIGINATING FROM THE CROSS: $A^b/lg\ a\ et \times lg\ a\ et$

| | SELFED PROGENIES | | | CROSSES WITH *a a* | | |
|---|---|---|---|---|---|---|
| PLANTS | PURPLE | PALE | COLORLESS | PURPLE | PALE | COLORLESS |
| 1A-1 | 0 | 125 | 48 | 0 | 788 | 866 |
| 8A-1 | 0 | 94 | 49 | 0 | 469 | 472 |
| 9A-1 | 0 | 68 | 26 | 0 | 148 | 144 |
| 14A-1 | 0 | 146 | 46 | 0 | 434 | 465 |
| 19A-1 | 0 | 40 | 14 | 0 | 693 | 695 |
| 78-1 | 0 | 76 | 18 | 0 | 143 | 159 |
| 78H-1 | .. | ... | .. | 0 | 360 | 365 |

Some of the *lg a et* plants used as a source of pollen in the 1947 crosses carried *Dt*. This gene induces the mutation of *a* to *A* and to other alleles of higher level. Rhoades[8,9] has shown that the frequency of this change is conditioned by the number of doses of both the *a* and *Dt* genes. The *a a a Dt dt dt* endosperms of seeds resulting from crosses with pollen which transmitted *Dt* showed a moderate degree of dotting on a colorless background. The cells of the purple endosperms ($A^b\ A^b\ a\ Dt\ dt\ dt$) carry both *a* and *Dt* and presumably some of these carry an *A* gene resulting from the *Dt*-induced mutation of *a*. However, no dots have been observed on purple endosperms of this constitution and it may be inferred that the mutant cells ($A^b\ A^b\ A$) are indistinguishable from $A^b\ A^b\ a$ cells. This failure of the mutated *A* to modify the phenotype of cells which already carry $A^b$ has an application in establishing with certainty the mutant character of the pale seeds suspected of carrying an $A^d$ allele. If the pale phenotype of these seeds is owing to the action of modifier genes and not to a change in the $A^b$ gene they as well as any pale seeds in their progeny are expected to be without dots. If, however, a mutation of $A^b$ to $A^d$ is involved, such that the endosperms are of the constitution $A^d\ A^d\ a\ Dt\ dt\ dt$ the cells carrying the *Dt*-induced, mutant $A$ gene ($A^d\ A^d\ A$) are expected to appear as purple dots on a pale background. Three of the seven pale seeds whose progeny test are given in table 1 had dotted endosperms. Two other seeds, though they had no dots, carried *Dt*. The absence of dots on these seeds may be attributed to the fact that dots frequently are absent on seeds which carry single doses of *Dt* and *a*. Each of the selfed progenies from these five seeds produced many pale, dotted endosperms. In a later generation all seven mutant cases were tested for the presence of dots on pale background by crossing the $A^d\ a$ plants with *a a Dt Dt* individuals. In all cases the presence of pale seeds with dots confirms the changed nature of the $A^d$ gene.

In view of the fact that the $A^d$ mutants originated in $A^b\ a$ plants it is necessary to rule out the possibility that they were derived from *a*. In a previous study[5] it was established that a number of mutants of the $A^d$ type which could have come only from $A^b$ share several distinctive types of effect: (1) These mutants, like the parent $A^b$ allele, are dominant to $A$ in their brown pericarp effect; (2) though the mutants are associated with a decidedly intermediate plant and aleurone phenotype, there is no cumulative effect with increasing doses; (3) in compounds with an allele of a higher level of action the $A^d$ mutants detract from, rather than add to, the effect of the former. Since mutants having this combination of effects have never been obtained from *a*, a study of the seven $A^d$ mutants under consideration may be expected to establish conclusively their origin. The original seven plants having the constitution $A^d\ a$ were crossed with $A\ A$ individuals. The pericarp color of the $F_1$ plants indicated, for each of the seven cases, that the $A^d$ gene is dominant to $A$ in its effect on this tissue. To test the effects of the $A^d$ mutants in compounds with stronger acting alleles the $A^d\ a$ plants were crossed with $A^{br}\ A^{br}$ individuals. In the $F_1$ progenies plants having the constitutions $A^{br}\ A^d$ and $A^{br}\ a$ were compared. Five of the seven mutants have been tested in this manner and all show a striking effect in the displacement of the phenotype of $A^{br}\ A^d$ plants in a direction away from wild type as compared with sib $A^{br}\ a$ plants. The tests of the dosage effects of the $A^d$ mutants require an additional generation and these studies are not completed. However, on the ears of the selfed $A^d\ a$ plants pale endosperms of three types occur, having one, two and three doses of the $A^d$ gene. The phenotypes of the pale seeds on these ears are quite uniform suggesting that the $A^d$ mutants show no dosage effect. More critical evidence on this point is being sought at the present time from studies involving comparisons in which a deficiency, instead of *a* is employed. These tests indicate that the $A^d$ mutants which were obtained in the present experiment from $A^b$ *a* individuals are indistinguishable from those which have a known origin from $A^b$ $A^b$ plants. It may be concluded that the former also are derivatives of $A^b$.

Since the heterozygous egg parents of the original cross were $A^b$/*lg a et* and the pollen parents were *lg a et*, a linkage of *lg* or *et* with $A^b$ in individuals from this cross is attributable to crossing over between $A^b$ and the loci of the marker genes. Table 2 summarizes the information on recombination for the *lg-a* and *a-et* regions obtained from progenies of the original crosses which produced the mutant individuals. Table 3 summarizes information as determined in progeny tests on the linkage phase of the etched endosperm and liguleless characters for the seven original plants carrying the mutant $A^d$ gene. The interesting feature of these data is that in six of the seven cases the chromosome carrying the $A^d$ gene emerged from the original cross combined with *et*.

TABLE 2

SUMMARY OF DATA ON CROSSING OVER IN THE *lg-a* (REGION I) AND *a-et* (REGION II) SEGMENTS FROM PROGENIES OF BACKCROSSES OF THE TYPE: $A^b$/*lg a et* × *lg a et**

| | CROSSOVERS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | REGION I | | REGION II | | REGIONS I AND II | | | |
| TOTALS | NUMBER | FREQUENCY | NUMBER | FREQUENCY | NUMBER | FREQUENCY | EXPECTED FREQUENCY | COINCIDENCE |
| 17,298 (51 ears) | 5714 | 0.3303 ±0.0036 | 2220 | 0.1283 ±0.0025 | 405 | 0.0234 ±0.0004 | 0.0424 | 0.55 |
| 22,706† (60 ears) | .. | | 2946 | 0.1297 ±0.0022 | | | | |

* The data reported are from the two families which were involved in the original crosses and which produced the seven individuals carrying the mutant $A^d$ gene.

† These more extensive data on crossing over in Region II include the 51 ears in the first row of the table along with additional progenies from the same families classified for *a* and *et* only.

TABLE 3

SUMMARY OF THE LINKAGE INFORMATION FOR THE MUTANT $A^d$-BEARING INDIVIDUALS FROM THE CROSS: $A^b$/*lg a et* × *lg a et*

| | CONSTITUTION OF THE $A^d$-BEARING STRAND | | | |
|---|---|---|---|---|
| | | CROSSOVER TYPES | | |
| MUTANT PLANT | NON-CROSSOVER TYPES | REGION I | REGION II | REGIONS I AND II |
| 1A-1 | ... | ..... | *Lg* $A^d$ *et* | ..... |
| 8A-1 | ... | ..... | " " " | ..... |
| 9A-1 | ... | ..... | " " " | ..... |
| 14A-1 | ... | ..... | " " " | ..... |
| 19A-1 | ... | ..... | " " " | ..... |
| 78-1 | ... | *lg* $A^d$ *Et* | ..... | ..... |
| 78H-1 | ... | ..... | ..... | *lg* $A^d$ *et* |
| Totals | 0 | 1 | 5 | 1 |

Unless the change of $A^b$ to $A^d$ is in some way related to crossing over it would be expected that the frequency of crossing over in the *a-et* interval would be the same in sporocytes which carry $A^d$ as in those which transmit $A^b$. On this basis, of the original mutant plants approximately 13% (table 2) $A^d$ *et/a et* (crossover) and 87% $A^d$ *Et/a et* (non-crossover) individuals are expected. The actual results were almost exactly the reverse of this; six of the seven individuals were crossover types for the *a-et* region. On the basis of 13% recombination and with no association between crossing over and mutation this number of mutants which are crossover types would be expected to occur among a total including approximately 40 mutants which are non-crossovers for the *a-et* region. There was only one of the latter.

The foregoing data suggest a relation between the mutation of $A^b$ and crossing over in the *a-et* region. This interpretation receives further support from experiments conducted in 1948 on a larger scale. As before, marked heterozygotes carrying $A^b$ were crossed with homozygous pollen parents of the constitution *lg a et*. In this case all of the pollen employed carried the *Dt* factor. Progeny tests of suspected mutant individuals from these crosses have not been made as yet. In the absence of these, however, all of the pale seeds with dotted endosperms have been selected and classified for normal versus etched phenotype. As indicated above by reason of their dotted phenotypes these seeds carry a mutant derivative of $A^b$ and their selection on this basis is not subject to the error and prejudice which may be involved when classification is made solely on the basis of a character which may be quite variable such as the degree of pigmentation. Of the total of 34 pale, dotted, mutant seeds, 30 are crossover types for the *a-et* interval. Here as in the previous experiment there is an unexpected preponderance of the crossover class. In order to determine the normal rate of crossing over in the stocks employed in this experiment the total progenies from those ears on which mutant seeds occurred were classified. Since the mutant seeds occurred singly on the ears, a total of 34 ears is involved. The value obtained (table 4) for recombination in the *a-et* region (13.9%) agrees reasonably well with that obtained in the 1947 experiment.

TABLE 4

SUMMARY OF DATA ON CROSSING OVER IN THE *lg-a* AND *a-et* SEGMENTS FROM BACKCROSSES OF THE MARKED HETEROZYGOTES CARRYING $A^b$ *

| | REGION I | | REGION II | | REGIONS I AND II | | | |
|---|---|---|---|---|---|---|---|---|
| TOTAL | NUMBER | FREQUENCY | NUMBER | FREQUENCY | NUMBER | FREQUENCY | EXPECTED FREQUENCY | COINCIDENCE |
| 5790 | 2068 | 0.3572 | 804 | 0.1389 | 184 | 0.0318 | 0.0496 | 0.64 |
| | | ±0.0063 | | ±0.0045 | | ±0.0023 | | |

* These data from 34 ears in 23 families in 1948.

The association of crossing over with mutation of $A^b$ may be explained if it is assumed that there is some causal relation between the two. On one hypothesis it may be argued that occasionally a crossover in the *a-et* region results in an unequal exchange of chromatin at the locus of a gene which, like $A^b$, affects the pigmentation of plant and endosperm tissues. On this basis the pale phenotype of the "mutant" individuals would not be due to a change in the $A^b$ gene but to a modification of chromatin somewhere between the *a* and *et* loci. Two types of evidence are against such an interpretation. (1) In all cases dots, resulting from the *Dt*-induced mutation of *a* to *A*, have been observed on the pale background of the mutant seeds or of pale seeds in their progenies. If the pale background of these seeds is owing to the action of a modifier gene, the phenotype of cells included in the dots should be modified to pale also. (2) If the pale phenotype of mutant seeds is held due to a modifier it is clear that the latter must be dominant since the cells of these endosperms carry a normal chromosome from the pollen parent. In this case the selfed or suitably crossed progenies of mutant individuals should produce some offspring which are homozygous recessive for the modifier condition and carry $A^b$; these would appear as recovered cases having purple endosperms and purple plant phenotypes. No such recoveries have been noted in the seven cases for which progenies are presently available.

TABLE 5

ACTUAL AND THEORETICAL FREQUENCIES OF CROSSOVERS AND NON-CROSSOVERS RELATIVE TO THE *a-et* INTERVAL FOR THE $A^d$-BEARING STRANDS OF THE FORTY-ONE MUTANT INDIVIDUALS*

| BASIS ON WHICH FREQUENCIES ARE COMPUTED | CROSSOVERS | NON-CROSSOVERS | STANDARD ERROR |
|---|---|---|---|
| Observed | 0.878 | 0.122 | 0.051 |
| No relation between mutation of $A^b$ and crossing over | 0.132 | 0.868 | 0.053 |
| Mutation of $A^b$ conditioned by crossing over at that locus. No allowance for interference | 0.868 | 0.132 | 0.053 |
| Same as above but allowing for coincidence of 0.5 | 0.934 | 0.066 | 0.039 |

* Actual values are based on the total of 41 mutants, 36 of which were crossover types. Theoretical values are based on the average rate of crossing over for this region (13.2%) taken from tables 2 and 4.

These two types of evidence not only rule out the possibility of a modifier effect in the occurrence of the mutant seeds but indicate further that the pale effect of the mutation must be due to a change in the $A^b$ gene itself. On any other basis the mutations of *a* (for which these seeds are heterozygous) to *A* through the action of *Dt* would not be recognizable as dots on these seeds; nor would the absence of purple seeds and plants in the progenies of the mutant individuals be expected. Since it is necessary to relate the high frequency of crossing over to this change the simplest and most direct hypothesis is that these mutations of $A^b$ to $A^d$ are a result of crossing over within the locus of $A^b$.

In the cases of 5 of the total of 41 mutants from both experiments the $A^d$-bearing strand delivered to the egg was seemingly a non-crossover type for the *a-et* interval. In view of the relatively great length of this segment (13.2 units as an average in these experiments) it is possible to explain these apparent non-crossover types as due to double exchanges, one occurring at the locus of $A^b$, giving rise to the mutation, the other occurring somewhere between this locus and that of *et*, thus reconstituting the parental combination. The data of table 5 are pertinent to this argument. Here are presented the actual frequencies of crossover and apparent non-crossover strands carrying the mutant $A^d$ gene alone with frequencies expected on three theoretical bases. The greatest discrepancy is between observed values and those expected on the basis of no relation between mutation and crossing over (second row). In the third row of the table are given the frequencies expected on the hypothesis that the mutation of $A^b$ is caused by a crossover at that locus. On this basis the non-crossover class results from double exchanges. Here there is good agreement between the theoretical and experimental frequencies. The value 0.122 for apparent non-crossovers among the 41 mutant cases is close to the average value of 0.132 for crossing over in the *a-et* region (from tables 2 and 4). However, in the computation of the theoretical frequencies in the third row of table 5 no allowance has been made for interference. The frequency of 0.122 for non-crossovers (reasoned as cases of double exchange) among the 41 mutant cases corresponds to a coincidence of 0.9. This exceeds the coincidence value of 0.6 corresponding to the *lg-a* and *a-et* regions (tables 2 and 4). In view of the small number of mutant cases it is not clear whether the excess of apparent non-crossover types among them is due to errors in sampling or to the occasional occurrence of mutations of $A^b$ which are not associated with crossing over. If mutations of the latter type occur a large proportion of these would fall in the non-crossover class thus tending to increase unduly the coincidence value. There is at present no basis for estimating the effect of a crossover at the $A^b$ locus on the frequency of coincidental exchange in the *a-et* region but it is reasonable to expect that the value for coincidence in this case would fall below 0.6. In the last row of table 5 are given the theoretical frequencies of crossovers and apparent non-crossovers associated with an arbitrarily chosen coincidence of 0.5; from the corresponding standard error for $N = 41$ it is clear that the actual frequencies could well have been chosen from such a population.

Confirmatory evidence of the association of mutation of $A^b$ and crossing over is being sought from independent experiments using plants in which the $A^b$ segment of the third chromosome either has no homologous region with which to synapse (in hemizygous plants) or, in cases where $A^b$ is carried in a duplication, competes at a disadvantage for synapsis. Since these experiments are designed to eliminate or suppress those $A^b$ mutations due to crossing over they may be expected to establish also whether mutations of $A^b$ to $A^d$ which are not associated with crossing over can occur.

In the foregoing presentation occasional reference has been made to the "$A^b$ locus." There is strong argument from the present evidence for viewing the "$A^b$ locus" as composed of at least two, physically distinct units which function coöperatively in the determination of the $A^b$ effect. If these are designated alpha and beta, the latter being more distal, the $A^d$ mutants may be described as having alpha and lacking beta since the strands carrying $A^d$ in these experiments are predominantly of the non-parental class for *et*, the more distal marker gene. This consideration poses several interesting problems whose analysis may be viewed as critical tests of the hypothesis. (1) Since on this basis $A^b$ is composed in part of the $A^d$ or alpha component the action of $A^b$ may be interpretable, at least in part, in terms of effects which, among those of the $A$ alleles, are distinctive for $A^d$. As reported elsewhere[5] these include the determination of brown pericarp which is dominant to the red effect of $A$, the absence of a cumulative effect with increasing doses, and a striking tendency to reduce rather than enhance the type effect in combinations with various $A$ alleles. Of the alleles associated with purple aleurone and plant phenotypes, only $A^b$ has a brown pericarp effect which is dominant; it is reasonable to infer that this action may be assigned to $A^d$ which has been argued as a component of $A^b$. It is not known whether $A^b$ shares the dosage and competitive effects of $A^d$ but this information is technically more difficult to obtain for $A^b$ than for $A^d$ because of the closer approach of the former to wild type. (2) If, as on the present scheme, the $A^d$ mutants represent the alpha component which has been separated from the beta constituent by crossing over between the two, it is expected that another class of mutants corresponding to the beta component carried on the reciprocal strands involved in such exchanges should occur with equal frequency. While these have not been isolated this is no argument against their occurrence. $A^b$ is associated with purple aleurone and plant while what has been argued as its alpha component, $A^d$, is distinctly intermediate in effect. From this it seems likely that the beta component produces a purple phenotype and if so mutants of this type would be indistinguishable from seeds carrying the unchanged $A^b$. However, on the expectation that the beta component produces a pericarp color distinguishable from $A^b$, experiments are being conducted to determine whether $A^b$ gives rise to mutants of this type. These would have escaped detection in the experiments reported here since the $P$ factor which has a complementary effect in the pigmentation of this tissue was absent. (3) Since the $A^d$ mutants are argued to represent a more proximal component of $A^b$ considerable interest attaches to whether these mutants are similar in their action. If the alpha component is itself compound in nature it might be predicted that the $A^d$ mutants in turn are not a homogeneous lot. Comparisons in constant genetic backgrounds of the mutant alleles of the present experiments are not available now; from observations of related but non-sib mutant plants the impression is given that these alleles will prove to be quite similar in effect. Nevertheless, the possibility of the component nature of alpha remains a particularly inviting prospect. Studies[5] on the action of $A^b$ and several mutants of the $A^d$ type indicate that their action cannot be explained on the basis of a simple relation between gene and reaction. Provisional schemes were offered to account for the exceptional behavior of these alleles. It was suggested (*a*) that the determinants corresponding to these genes function as units but that the enzyme corresponding in each case might participate in more than a single reaction, or alternatively (*b*) that a simple relation biochemically between gene and enzyme might be defended on the assumption that each of these alleles is compound in the sense of being constituted by more than one physically distinct determinant. The evidence reported here suggests that the antimorphic action of $A^b$ resides in such a component nature. It is possible that the $A^d$ mutants likewise owe their unusual behavior to a similar cause but proof of this requires direct evidence of the compound nature of what has been labeled here the alpha component. (4) Is the crossover which gives rise to the mutants preceded by unequal synapsis? Since the mutations occurred in plants heterozygous for $a$ it follows that any statement about such an arrangement must be based on some understanding of the physical constitution of this allele and on this point there is no information. There is hence no basis from these experiments to argue that the change is related to unequal synapsis or even that there is any final inequality of exchange. However the mutants which Stadler[6] obtained and which are similar in effect to the $A^d$ mutants of the present experiments occurred in homozygous $A^b$ individuals. It is not known whether these mutations were associated with crossing over, but if so, there is a basis for the argument that unequal synapsis is involved. Making use of more efficient markers which have become available recently, the relation between crossing over and mutation in $A^b A^b$ plants is being investigated.

It is premature to discuss the relation of the present case to others which have been more extensively studied.[1–9] Nevertheless it is a rather inviting prospect, in analogy with the cases of Bar, lozenge, Star-asteroid and others[10] in *Drosophila* and with that in *Gossypium*,[11, 12] that $A^b$ represents repeated loci and by further analogy with several of these cases has components which are derived by duplication from a common ancestral locus and which have since become modified. No less attractive is the possibility that the divergent action of certain of the *A* alleles may find an explanation in terms of their component nature.

---

* This work was conducted at and supported by Princeton University, California Institute of Technology and the University of Illinois, to which institutions grateful acknowledgment is made.

[1] Sturtevant, A. H., *Genetics*, **10**, 117–147 (1925).

[2] Oliver, C. P., *Proc. Nat. Acad. Sci.*, **26**, 452–454 (1940).

[3] Lewis, E. B., *Genetics*, **30**, 137–166 (1945).

[4] Emerson, R. A., and Anderson, E. G., *Ibid.*, **17**, 503–509 (1932).

[5] Laughnan, J. R., *Ibid.*, **33**, 488–517 (1948).

[6] Stadler, L. J., *Cold Spring Harbor Symp. Quant. Biol.*, **9**, 168–177 (1941).

[7] Stadler, L. J., and Laughnan, J. R., unpublished.

[8] Rhoades, M. M., *Genetics*, **23**, 377–397 (1938).

[9] Rhoades, M. M., *Cold Spring Harbor Symp. Quant. Biol.*, **9**, 138–144 (1941).

[10] Lewis, E. B., *Genetics* (abstract), **33**, 113 (1948).

[11] Yu, C. P., and Chang, T. S., *J. Genet.*, **49**, 46–56 (1948).

[12] Stephens, S. G., *Genetics*, **33**, 191–214 (1948).

---

# *ABERRANT HETEROZYGOTES IN ESCHERICHIA COLI**

BY JOSHUA LEDERBERG

DEPARTMENT OF GENETICS, UNIVERSITY OF WISCONSIN

Communicated by R. A. Brink, February 9, 1949

A mechanism of genetic recombination has been indicated in experiments on *Escherichia coli*, strain K-12.[1] A synthetic agar medium was used as a selective sieve to isolate occasional prototroph recombinants which appear in mixed cultures of complementary biochemical mutants. Later, additional genetic factors were introduced, including fermentation and virus-resistance mutations, and these factors were found to segregate in characteristic ratios, suggesting linkage. All the cells in a given prototroph colony showed the same combination of characters and were stable on further cultivation. Therefore, it was inferred that segregation had occurred before the initiation of the colony, and that the postulated zygote had a very short life, probably a single cell generation. The life cycle would resemble the ascomycete's, in which haploid nuclei fuse to form a transient diploid zygote which undergoes meiosis without any intervening mitoses.

Exceptions to this rule have now been found in the form of unstable prototroph cultures which continually segregate out various recombination types so as to suggest that they are heterozygous and diploid. But abnormalities in the segregation ratios suggest some sort of chromosomal aberration whose nature has not yet been proved.

*Material and Methods.*—The cultural and crossing techniques used earlier have generally been followed. Crosses were also made by spreading mixtures of cells on the surface of synthetic medium[2] to which carbohydrate and an eosin-methylene blue indicator are added. This medium, EMS-, classifies prototrophs as soon as they appear for fermentation of the sugar used. Other media used include EMB- agar, in which the indicator is incorporated in a sugar peptone base. The genetic factors involved in this study are summarized in table 1.

TABLE 1

SUMMARY OF MUTANTS

| LOCUS | "TYPE" ALLELE | MUTANT ALLELE | CHARACTER OF MUTANT |
|---|---|---|---|
| B | + | − | Needs biotin for growth |
| M | + | − | Needs methionine |
| T | + | − | Needs threonine |
| L | + | − | Needs leucine |
| $B_1$ | + | − | Needs thiamin |
| Lac | + | − | Unable to ferment lactose |
| Mal | + | − | Unable to ferment maltose |
| Gal | + | − | Unable to ferment galactose |
| Ara | + | − | Unable to ferment L-arabinose |
| Xyl | + | − | Unable to ferment D-xylose |
| Mtl | + | − | Unable to ferment mannitol |
| $V_1$ | *s* | *r* | Resistant to phages T1, T5 |
| $V_{1c}$ | *s* | *r* | Resistant to T1; sensitive to T5 |
| "*Het*" | .. | .. | Persistent heterozygote factor |

*Crosses Involving "Het."*—The first unstable prototroph to be discovered, "*H-1*," arose in a cross to test the allelism of the virus resistance factors $V_1$ and $V_{1c}$. The parents were B − M − Lac + $V_{1c}^r$ and T − L − $B_1$ − Lac − $V_1^r$. Among 200 prototrophs picked and tested on synthetic agar for sensitivity to T1, only one was an apparent crossover showing a full sensitive reaction. However, when this unique culture was streaked out on EMB − lactose, it gave rise not to typical dark Lac+ or light Lac− colonies, but to highly variegated colonies with intermingled sectors of light and dark cells. The variegated reaction will be referred to as Lac*v*.

When Lac*v* colonies were streaked out again on EMB − lactose, pure Lac−, pure Lac+ and a few Lac*v* colonies were seen. But when the pure − and + colonies were inoculated on EMS − lactose they failed to grow, showing that they were nutritionally deficient. Furthermore, although they had been derived from a virus-sensitive culture, most of the purified "segregants" were resistant to T1 when tested on EMB − lactose. On the other hand, Lac*v* colonies gave rise to normal appearing Lac+ on

EMS—lactose, which, as before were sensitive to T1. Taken back to EMB—lactose, they showed the Lac*v* result of splitting off pure − and + types. When single colonies were streaked out in series for 15 transfers on EMS—lactose, they retained their normal Lac+ appearance, but throughout the experiment they continued to segregate when brought on EMB—lactose, which is a nutritionally "complete" medium. To summarize, *H-1* was an unstable prototroph culture which was apparently Lac+ and T1-sensitive so long as it was maintained on a synthetic medium, but which segregated nutritionally deficient, Lac− and T1-resistants on a "complete" medium.

At this point, three alternatives had to be distinguished. The segregating entities might be (*a*) associations of intact cells of the original parental mutants collaborating by syntrophism to allow growth on synthetic medium; or (*b*) heterokaryons comprising intact nuclei of the parents in the same cell; or finally (*c*) diploid heterozygotes. The phage sensitivity of *H-1*, contrasted to the resistance of the parents, render (*a*) unlikely, but the segregation might still be from a heterokaryon or a heterozygote. To test these alternatives, it was thought that a heterokaryon should split out each of the two parental genotypes, but no additional combinations. On the other hand, segregation from a heterozygote should sometimes be accompanied by crossing over, and the formation of new combinations of the parental characters.

One hundred thirty-five Lac*v* colonies were separately streaked out on EMB—lactose, and a single pure Lac+ and Lac− segregant isolated from each of the 135 streakings, to insure that each of the cultures tested was an independent segregant. Among the 135 Lac− tested, there were 121 T5$^r$ (parentals) and 14 T5$^s$ (exchanges); the Lac+ included 133 T5$^s$ (parentals) and 2 T5$^r$ (exchanges). Therefore, out of 270 tests there were 16, or 6%, new combinations. Therefore, crossing over occurs during the segregation of *H-1*. This is compatible only with the hypothesis of segregation from a heterozygote.

The following remarks should not obscure this major conclusion. The amount of exchange in the Lac− and Lac+ series is not the same ($\chi^2_1 = 9$). The crossovers in the two series came from different Lac*v* colonies, but reciprocality cannot be demanded if many different segregations occur during the growth of a single Lac*v* colony. Even the higher value of $^{14}/_{135}$ or 10% for exchanges between Lac and $V_1$ in the Lac− series is lower than the 38% derived in previous studies on prototroph recombinants. Finally, it was very obvious that the Lac− segregants far outnumbered the Lac+. If anything, selective pressures on EMB—lactose seem to favor Lac+, so this finding was most unexpected.

Owing to technical difficulties, tests on nutritional segregations have been less extensive. It proved to be especially difficult to score for the biotin requirement in the presence of amino acids, probably due to partial replacement. The classification of the thiamin requirement ($B_1-$) also requires fastidious attention to clean chemicals and glassware. Most of the Lac+T5$^s$ segregants are the parental B − M −; Lac − T5$^r$ are usually T − L − $B_1-$. However, prototrophs and the multiple mutant combinations M − T − L − and M − T − L − $B_1-$ have been encountered. These combinations could not be obtained by previous methods, which necessarily isolated prototrophs.

Unstable heterozygotes like *H-1* had not been encountered before, although many hundreds of Lac+ prototrophs from crosses heterozygous for Lac had been streaked out on EMB − lactose. The persistence of *H-1* as a diploid might be a non-genetic accident or perhaps an effect of a spontaneous life cycle mutation. If so, it would be possible to set up crosses to yield diploids heterozygous for additional markers.

Three segregants from *H-1* have been used in crosses to test for the production of persistent heterozygotes. (*A*) (W-466) was B − M − Lac$_1$ − $V_1^s$; (*B*) (W-477) T − L − $B_1$ − Lac$_1$ − $V_1^r$ and (*C*) (W-478) B − M − Lac + $V_1^s$. These cultures were crossed on EMS − lactose with appropriate, "standard" complementary stocks (B − M − or T − L − $B_1-$; Lac − or Lac+) of independent origin. Lac+ prototrophy were streaked out on EMB − lactose, and Lac*v* colonies looked for. In each of these crosses, about 5% of the Lac+ prototrophs were heterozygous. Crosses of *A* × *C* gave the same result. If "*Het*" is located on a chromosome, it is effective whether present in one or both parents.

In further studies still in progress, *C* has been crossed with multiple fermentation mutant stocks, T − L − $B_1$ − Lac$_1$ − Mal − Gal − Ara − Xyl −, and some also Mtl −. Such stocks were developed by irradiations on EMB − media, in sequence. As before, about 5 to 10% of the + prototrophs isolated from EMS lactose or xylose plates are heterozygous. However, these cultures are not uniformly heterozygous for all the factors in which the parents differed. It is especially notable that in over a hundred heterozygotes obtained in this way, Mal has never been heterozygous. Usually, the cultures are pure Mal −, sometimes Mal+. Xyl*v* cultures are equally often Lac*v* or pure for Lac, as well as the converse.

It remains to be decided whether the "pure" characters are homozygous or hemizygous. The Mal locus which, as mentioned, has always been "pure" is probably hemizygous, as shown by reversion studies. If the cultures were homozygous, Mal − would be twice represented, i.e., Mal − /Mal −. Reverse mutations of one of these genes should lead to heterozygosity at the locus. Preliminary tests on reversions for Mal in these stocks have invariably given pure Mal+ cultures, although they still segregated for other factors. This suggests that Mal is represented only once, that it is hemizygous, and, therefore, that there is deficiency for a region including this locus. This tentative conclusion is in accord with the aberrant segregation of Lac (excess of Lac−) observed in *II-1*. The hypothetical deficiency would be lethal in a haploid, and lead to the loss of segregants carrying the alleles linked to it.

Some quantitative studies of segregation and crossing over have been carried out on these complex heterozygotes. *II-72* segregates both for lactose and xylose fermentation. Segregation was permitted to occur "enmasse" in heavy cultures in nutrient broth, the cultures diluted and plated on EMB− xylose and EMB−lactose. Of 895 colonies on EMB−xylose, 19 were still Xyl*v*; 815 Xyl− and 61 Xyl+. The Xyl+ were all Lac−. The segregation ratio is 815−:61+ or more than 13:1. Similarly, of 753 colonies on EMB−lactose, 23 were still Lac*v*; 654 Lac− and 76 Lac+ (all of which were Xyl−). Thus, no Lac+Xyl+ segregants were observed in this sample. The other classes are computed to be Lac−Xyl− 83%, Lac−Xyl+ 7% and Lac+Xyl− 10%. These figures might be accounted for by mapping:

Xyl+ *7.7* Lethal *11.2* Lac+/Xyl− . . . Lac−.

To justify these conclusions, independent mapping of Xyl, Mal, Lac will be necessary.

Four hundred ten segregants were tested both on EMB−lactose and on EMB−xylose. Segregation was always complete; i.e., no cultures were found which had segregated for one factor, and not for the other. This eliminates interpretations based on new segmental losses, or on some forms of autogamy.

It may be argued that the unequal segregations are due to selection. This factor cannot be excluded, but the dominant segregant differs from diploid to diploid. That is to say, other heterozygotes have been isolated from parallel crosses which gave an excess of Lac+ rather than Lac− segregants, or Xyl+ more than Xyl− or both. It seems more likely that the ratios observed are based upon the genetic constitution of the heterozygote.

The Lac+ character of $\text{Lac}_1$ heterozygotes shows that $\text{Lac}_1$+ is dominant to $\text{Lac}_1$−. A number of other loci have been identified, mutations at which lead to the inability to ferment lactose. At two more of these at least, the + allele is dominant. Cultures carrying $\text{Lac}_2$− or $\text{Lac}_4$− were crossed to "*Het*" stocks carrying $\text{Lac}_1$−. A large proportion of the Lac+ prototrophs seen on EMS−lactose turned out to segregate for lactose fermentation. Their variegation, however, was generally periclinal rather than sectorial, with dark, Lac+ centers and light, Lac − margins with only occasional streaks of Lac+. This difference is to be expected, since each of the chromosomes of the diploid carries a Lac− mutation. Therefore, only those segregants in which there has been a crossover between the two Lac loci, bringing the + alleles into coupling phase will be lactose positive. That these cultures are segregating for two Lac− factors has been confirmed by physiological and genetic tests on the segregants.

*Heterozygotes from Standard Crosses.*—All the heterozygotes so far referred to are the issue of crosses involving the hypothetical "*Het*" factor derived from *H-1*. In these crosses, several per cent of the prototrophs are demonstrably heterozygous. Previous and current controls showed that, if they occurred at all in "normal" crosses, they must be very much rarer. In order to make a more thorough test, advantage was taken of the very close linkage (less than 1% recombination) which has been observed between the Lac$_1$ and Lac$_4$ loci mentioned above. B−M−Lac$_1$−Lac$_4$+ was crossed with T−L−B$_1$−Lac$_1$+Lac$_4$−. Much less than 1% of the prototrophs of this cross on EMS−lactose are Lac+. It was reasoned that a Lac+ prototroph might represent either a very rare crossover, or a diploid in which the + factors were carried on opposite chromosomes. Because most haploid recombinants, being Lac−, can be set aside by inspection, this is a fairly efficient way of screening for rare diploids. In this way, heterozygous diploids were also obtained from normal stocks, not carrying "*Het*," but they constitute only about 0.1% of the prototrophs. These prototrophs are generally similar to the previous ones, except that they are somewhat more stable. It has not been established whether new "*Het*" mutations have occurred here. When these diploids are singly heterozygous for various sugar fermentation factors, they also show non-random segregations, but these mutations are not the same as those in the "*Het*" series, so that a direct comparison is not yet possible.

*Discussion.*—It is not clear what relationship there is between the exceptional persistence of these diploids, and the abnormalities in their segregation. There may be some direct connection between their heteroploidy, i.e., deficiency for the Mal region, and the prolongation of the diplophase. But the very disturbing possibility has not been discounted that these heterozygotes merely reveal a situation which also operates in the formation of haploid prototrophs from transient zygotes. If this is so, the linkage map of *E. coli* K-12 will have to be systematically reexamined, with the use of several unrelated sets of stocks. In any case, the discrepancies affect only the details of chromosome behavior. These heterozygotes, on the other hand, have provided an unexpected confirmation of the sexual basis of genetic recombination in this bacterium.

Although the heteroploidy is somewhat of a limitation, heterozygote formation is a very helpful tool in genetic analysis. Since all types of recombinants, not only prototrophs, can be recovered, it is feasible to extend pedigrees to several generations, and interesting combinations of factors can be put together in a form allowing their use in crosses. For example, a $Lac_1-$ mutation occurring in a $T-L-B_1-$ stock has been transferred to a combination with $B-M-$, allowing it to be compared genetically in crosses with other $Lac-$ mutants induced in $T-L-B_1-$ material.

The experimental production and maintenance of diploids raises questions of dominance and dosage effects. It has already been noted that the normal alleles of several biochemical mutations are dominant, although quantitative comparisons of enzyme competence in heterozygotes, haploids and homozygotes remain to be carried out. The dominance of the type sensitive alleles of genes for phage resistance supports, at least in part, a segregation interpretation for the delayed effects noted by Demerec and Latarjet in induced resistance mutations.[3] On the other hand, for the study of phenotypic lag in bacteria,[4] segregation provides a far larger and more reproducible source of material than mutation.

In respect to lactose fermentation, such a lag cannot last more than a few cell generations in this material, in view of the appearance of numerous $Lac-$ sectors in Lac*v* colonies on $EMB-$lactose agar. The masking effects noted[5] with yeast asci segregating in the presence of the substrate have thus not beeen found here.

Cytological and single cell studies on these strains are in progress.

*Summary.*—1. Unstable prototrophs have been isolated from certain crosses in *E. coli* K-12, and characterized as segregating heterozygous diploids.

2. The capacity to produce appreciable numbers of persistent heterozygous diploids is inherited.

3. The segregation of various mutant factors is strongly biased, possibly due to a recessive lethal deletion including a locus affecting maltose fermentation.

4. In the heterozygotes, the type + alleles of factors controlling several fermentations and nutritional requirements are dominant. Sensitivity to bacteriophage T1 is dominant to resistance.

---

[*] Paper No. 391 from the Department of Genetics, College of Agriculture, University of Wisconsin. This work has been aided by a grant from the Wisconsin Alumni Research Foundation. The author is indebted to Professor James F. Crow for his stimulating discussions.

[1] Tatum, E. L., and Lederberg, J., *J. Bact.*, **53**, 673–684 (1947). Lederberg, J., *Genetics*, **32**, 505–525 (1947).

[2] The composition of EMS-, in grams per liter, is: Sodium Succinate 5; $(NH_4)_2SO_4$ 5; NaCl 1; $MgSO_4$ 1; $K_2HPO_4$ 2; Agar 15; Methylene Blue Hydrochloride 0.065; Eosin Y 0.4. This formula is conveniently stored as a dry mixture of the powdered components.

[3] Demerec, M., and Latarjet, R., *Cold Spring Harbor Symposia Quant. Biol.*, **11**, 38–50 (1946).

[4] Newcombe, H. B., *Genetics*, **33**, 447–476 (1948).

[5] Spiegelman, S., Lindegren, C. C., and Lindegren, G., these PROCEEDINGS, **31**, 95–102 (1945).

# THE IN VITRO SYNTHESIS OF PANTOTHENIC ACID BY PANTOTHENICLESS AND WILD TYPE NEUROSPORA

BY ROBERT P. WAGNER*

GENETICS LABORATORY, UNIVERSITY OF TEXAS, AUSTIN, TEXAS

Communicated by J. T. Patterson, February 26, 1949

It has been demonstrated by Wagner and Guirard[1] that the resting mycelium of wild type *Neurospora* is able to synthesize pantothenic acid in the presence of $\beta$-alanine and pantoyl lactone (DL-$\alpha$-hydroxy-$\beta$,$\beta$-dimethyl-$\gamma$-butyrolactone). The *in vivo* synthesis of pantothenate from these precursors cannot be performed, however, by the pantothenicless mutant 5531. If the wild type mycelium is treated with acetone to kill the cells and then dried with acetone, the resultant preparation can also be shown to possess an active enzyme system capable of carrying out the above synthesis *in vitro*.[1]

This present communication is concerned with an extension of the investigation of *in vitro* synthesis of pantothenate by *Neurospora*. Three strains, Emerson wild type 5256A, and two pantothenicless mutants, 5531A and 34556A, were used. Both mutants appear to be due to a change at the same gene locus. The two differ in that 5531 was induced by x-rays and 34556 by ultra-violet light.[2]

*Experimental.—Activity of Intact and Water Extracted, Dried Mycelium:* Fresh Mycelium of 5256, 34556, and 5531 was treated with acetone according to the method previously described.[1] The resultant preparations are described in what follows as intact, acetone-dried mycelium. The activity of these preparations with respect to the production of pantothenic acid from the precursors, $\beta$-alanine and pantoyl lactone was tested at three different temperatures. In each case the reaction mixture consisted of 50 mg. of acetone-dried material/flask suspended in 25 ml. of 0.04 $M$ potassium phosphate buffer at pH 6.5. In the case of those flasks to which precursors were added, an initial concentration of 0.004 $M$ of each precursor was used. The flasks and contents were sterilized by autoclaving and cooled before the addition of the acetone preparations. Incubation was carried on for twenty-four hours with constant shaking at 22°, 27° and 35°C. The flasks were then steamed, the solids filtered out and the filtrate assayed for pantothenate with *Lactobacillus arabinosus* by the method of Skeggs and Wright.[3] The results are presented in table 1, as $\mu$g. pantothenate produced/50 mg. mycelium/twenty-four hours, both in the presence and absence of the two precursors.

Ten grams of each of the intact, acetone-dried preparations were washed in 1000 ml. of distilled water for several hours with constant shaking and then filtered. The residues (R) were dried with acetone and weighed.

In the case of each strain the residue was found to have lost ca. 50% of the original weight of the unextracted preparation. The filtrates were each reduced *in vacuo* to 200 ml. The water-insoluble residues (R) were tested for activity in the same way that the intact preparations were tested using 50 mg. of R per flask. The concentrated filtrates (S) were sterilized by Seitz filtration and tested for activity at 35° with and without the corresponding residues. An amount of S (1 ml.) equivalent to 50 mg. of intact, acetone mycelium was added to the flasks. Table 2 presents the results of these tests.

*Activity of the Concentrated Water Extract:* The concentrated filtrate (S) from each strain was added in several different ml. concentrations to a set of flasks containing the precursors and 50 mg. of the respective residues. Incubation was at 35° for twenty-four hours. The results are described by figure 1.

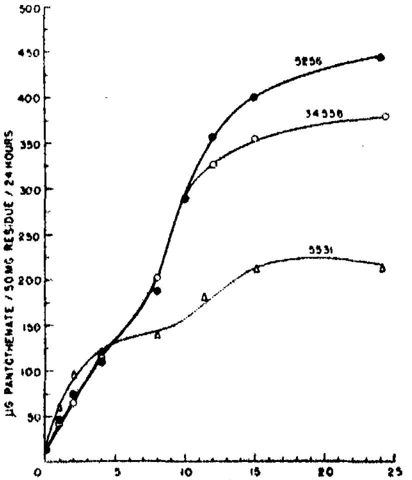


FIGURE 1

The effect of added excess filtrate (S) on the production of pantothenate.

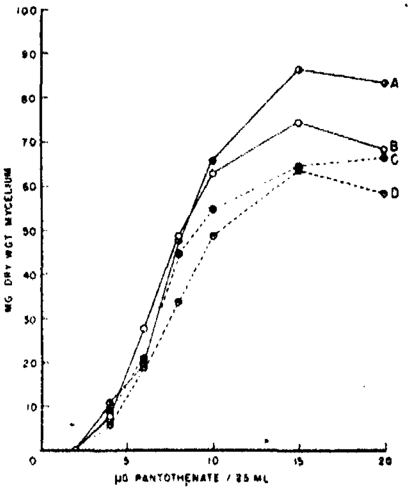


FIGURE 2

The effect of the precursors on the growth of the pantothenicless mutants. A, 5531 plus precursors; B, 5531 without precursors; C, 34556 without precursors, D, 34556 plus precursors.

*The Effect of the Precursors on the Growth of the Mutants:* The mutants 34556 and 5531 were grown for four days at 35° and pH 6.5 in the presence of 0.004 *M* β-alanine and pantoyl lactone, and various concentrations of Ca pantothenate. A control series without the precursors, but with Ca pantothenate, was tested simultaneously. The results of this experiment are given in figure 2.

*Growth of the Mutants on Pantothenate Synthesized in Vitro:* The pantothenicless mutants were inoculated into media containing various concentrations of pantothenate synthesized by the acetone preparations of these same strains which had been previously assayed with *L. arabinosus*. The pH of the medium was at 6.5, the temperature, 35°, and the incubation period, four days. Controls were run simultaneously with Ca pantothenate. Figure 3 shows the growth response of 5531 to pantothenate synthesized by 5531 *in vitro*, and figure 4 the response of 34556 to its corresponding source of pantothenate. The data from the corresponding controls are included in each case for comparison.

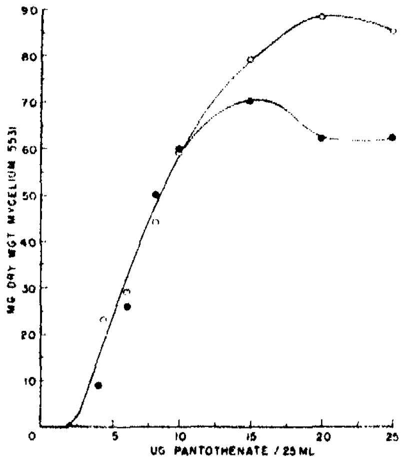


FIGURE 3

Growth of 5531 on Pantothenate synthesized *in vitro* by 5531.

O pantothenate synthesized by 5513

● Ca pantothenate control

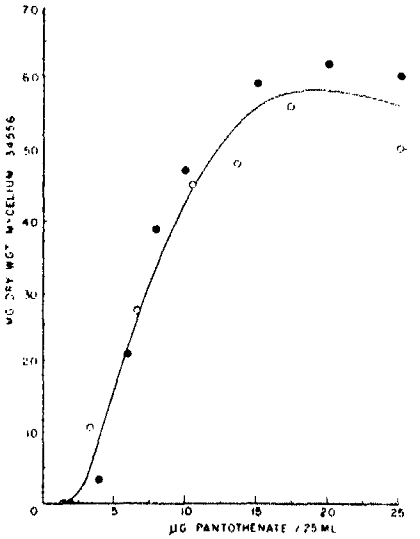


FIGURE 4

Growth of 34556 on pantothenate synthesized *in vitro* by 34556.

O pantothenate synthesized by 34556

● Ca pantothenate control

*Discussion.*—The data presented in tables 1 and 2 demonstrate that the pantothenicless mutants, 5531 and 34556, as well as the wild type *Neurospora*, possess an enzyme system capable of carrying out the *in vitro* synthesis of pantothenic acid from β-alanine and pantoyl lactone. There is a difference indicated in the system possessed by the mutants when compared to the wild type with regard to the effect of temperature, but the enzyme is active to the same degree in all three strains at 35°. The enzyme in each strain consists of at least two parts, a water-insoluble residue (R) and a water-soluble portion (S) which is inactive alone, but which increases the activity of R when combined with it.

The data in figure 1 when combined with the data in tables 1 and 2 show that the concentration of S is limiting, with the concentration of precursor used, in the synthesis of pantothenate by the intact and washed acetone preparations. The amount of pantothenate synthesized in the presence of excess S, as shown in figure 1, leaves no doubt that there is present in all three strains a mechanism active in this synthesis. This considerable increase in pantothenate production cannot be accounted for by the pantothenate present in S, which is 0.31, 0.48 and 0.65 μg./ml. for 5256 (S), 5531 (S) and 34556 (S), respectively.

TABLE 1

THE ACTIVITY OF INTACT, ACETONE-DRIED MYCELIUM

| ACETONE PREPARATION | 22° | | 27° | | 35° | |
|---|---|---|---|---|---|---|
| | O | P | O | P | O | P |
| 5256 | 4.0 | 12.5 | 2.8 | 55.0 | 2.3 | 48.0 |
| 5531 | 0.51 | 0.55 | 1.8 | 13.0 | 1.5 | 47.0 |
| 34556 | 4.2 | 8.3 | 1.5 | 9.5 | 1.5 | 45.0 |

μg. pantothenate produced/50 mg. intact, acetone-dried mycelium/twenty-four hours by 5256 (wild type) and two pantothenicless strains, 5531 and 34556, at three different temperatures. O designates absence and P presence of both pantoyl lactone and β-alanine.

TABLE 2

THE ACTIVITY OF WATER-SOLUBLE AND WATER-INSOLUBLE FRACTIONS

| REACTION MIXTURE | 5256 | 5531 | 34556 |
|---|---|---|---|
| Residue alone | 0.43 | 0.55 | 0.90 |
| Residue + precursors | 12.0 | 13.0 | 12.0 |
| Filtrate + precursors | 0.42 | 0.65 | 0.65 |
| Residue + filtrate + precursors | 48.0 | 78.0 | 68.0 |

μg. pantothenate produced/50 mg. residue/twenty-four hours. The amount of filtrate added was equivalent to 50 mg. of intact acetone dried mycelium. Temperature at 35°C.

The water-soluble factors appear to be identical in activity for the three strains, since they can be substituted for one another with results comparable to those given in figure 1. In addition the factor or factors are dialyzable and heat stable while the water-insoluble fraction is not. Further work has been done on the coenzyme indicated, and these results will be recorded elsewhere. It is evident that 34556, 5531 and 5256 each have the same or a similar coenzyme involved in the production of pantothenic acid from β-alanine and pantoyl lactone.

In all of the experiments discussed above the production of pantothenate was determined by assaying with *L. arabinosus.* It was necessary, therefore, to establish the activity of the pantothenate for the mutants. Since the reaction mixtures containing pantothenate also had uncombined pantoyl lactone and β-alanine present, the effect of these compounds on the growth of the mutants was determined as shown in figure 2. It can be seen from these data that no growth occurs in the absence of pantothenate with the precursors present, and there is no increase in growth over the controls in the presence of the precursors and limiting amounts of pantothenate. There does seem to be a slight but consistent beneficial effect of the precursors on the growth of 5531 at optimal concentrations of pantothenate.

In view of the inactivity of the precursors, the data in Figures 3 and 4 definitely establish the activity of the pantothenate synthesized by 5531 and 34556 *in vitro* for these mutants. The results reported here were obtained by using pantothenate synthesized in the presence of excess coenzyme. Similar results were obtained by using the pantothenate synthesized by intact, acetone-dried mycelium, and also with pantothenate synthesized by wild type.

These findings are of interest in connection with the current hypotheses dealing with the postulated gene enzyme relationship. According to one of these hypotheses[4] 5531 and 34556 are mutant because they are incapable of synthesizing their own pantothenic acid due to a blocking of the reaction coupling $\beta$-alanine and pantoyl lactone. The block may be caused by the absence of the necessary enzyme or coenzyme or by inhibition. This interpretation is supported by the fact that (1) wild type is stimulated to produce pantothenic acid by the presence of these two compounds,[1] (2) by the demonstration that these compounds together do not support growth of the mutants, and (3) by the report that these precursors accumulate in the medium when the mutant 5531 is grown in the presence of pantothenate.[5] (Strain 34556 has not as yet been tested for accumulation of precursors.) The data presented in this report rule out the enzyme absence hypothesis for these mutants, assuming that we are correct in concluding that it is this particular step in the synthesis of pantothenate which is involved. It is then necessary to conclude that the mutants possess a mechanism which is active *in vivo* in preventing this enzyme system from operating in pantothenate synthesis.

*Summary*.—The pantothenicless mutants 5531 and 34556 of *Neurospora* possess, like wild type, an enzyme system for the synthesis of pantothenic acid from $\beta$-alanine and pantoyl lactone *in vitro*.

---

[*] A portion of this work was done while the author was a research fellow at the California Institute of Technology during the summer of 1948.

[1] Wagner, R. P., and Guirard, B. M., these PROCEEDINGS, **34**, 398–402 (1948).

[2] Houlahan, M., Private Communication.

[3] Skeggs, H. R., and Wright, L. D., *J. Biol. Chem.*, **156**, 21 (1944).

[4] Beadle, G. W., *Chem. Rev.*, **37**, 15–96 (1945).

[5] Tatum, E. L., and Beadle, G. W., *Ann. Missouri Botan. Garden*, **32**, 125–29 (1945).

# BOUNDS FOR THE SOLUTIONS OF A SECOND-ORDER LINEAR DIFFERENTIAL EQUATION

By Walter Leighton

Department of Mathematics, Washington University

Communicated by Marston Morse, February 15, 1949

In this note we consider self-adjoint differential equations of the form

$$(r(x)y')' + p(x)y = 0, \tag{1}$$

where $r$, $p$ and $(rp)'$ are continuous and $r$ and $p$ are positive on $I$: $a \leq x < \infty$.

Theorem. *If $(rp)' \geq 0$ on $I$, every solution $y(x)$ of (1) is bounded on $I$. If $(rp)' \leq 0$ on $I$, $|y'(x)| \leq M/r(x)$ on $I$.*

It is easy to verify that if $y(x)$ is a solution of (1),

$$(ry')^2 + rpy^2 = c^2 + \int_a^x (rp)'y^2dx, \tag{2}$$

where $c^2$ is a positive constant. Suppose the maximum $M$ of $|y(x)|$ on $[a, x]$ is assumed at $x = \xi$. Then

$$r(\xi)y'(\xi)^2 + r(\xi)p(\xi)M^2 = c^2 + \int_a^\xi (rp)'y^2dx,$$

and

$$r(\xi)p(\xi)M^2 \leq c^2 + M^2 \int_a^\xi (rp)'dx.$$

It follows that $M^2 \leq c^2/r(a)p(a)$. Since $c^2/r(a)p(a)$ is independent of $x$, the first statement in the theorem is proved. The second statement in the theorem is an obvious consequence of (2). One notes incidentally that if $(rp)' \leq 0$ on $I$, $rpy^2 \leq M$.

In the corollaries which follow the theorem is applied to determine bounds for the solutions of a differential equation in so-called *normal* form of (1):

$$y_0'' + p_0(x)y_0 = 0, \tag{3}$$

where we assume $p_0(x)$ to be class $C'$ on $I$. It will be observed that the method employed is also available for equations not in this form.

Corollary 1. *If $R(x)$ is a positive function of class $C'''$ on $I$ such that*

$$rp = R^3(R'' + p_0R) > 0 \tag{4}$$

*and $(rp)' \geq 0$ on $I$, every solution $y_0(x)$ of (3) satisfies the condition*

$$|y_0(x)| \leq MR(x).$$

To prove the corollary it is sufficient to note that the transformation $y_0 = r^{1/2}y(r^{1/2} = R)$ carries (3) into (1), where $p$ is defined by (4).

"Bestness" of the corollary's conditions follows from the readily verifiable fact that the general solution of the differential equation $R^3(R'' + p_0R) = 1$ can be given the form $R^2 = ay_1^2 + by_2^2 + 2cy_1y_2$ $(ab = \delta^{-2} + c^2)$, where $\delta$ is the Wronskian of the solutions $y_1$ and $y_2$ of (3).

A frequently useful consequence of Corollary 1 may be obtained by writing $x^2p_0 = q$. We have then the following result.

COROLLARY 2. *If there exists a number* $\alpha$ *such that* $\pi = x^{4\alpha-2}(q - \alpha + \alpha^2) > 0$ *and* $\pi' \geq 0$ *on* $I$, *a solution* $y_0(x)$ *of* (3) *has the property that* $|y_0(x)| \leq Mx^{\alpha}$ *on* $I$.

To prove this corollary one sets $R = x^{\alpha}$ in Corollary 1.

The method suggested in the preceding corollaries is also available for determining bounds for the derivative of a solution $y(x)$ of (1). One simply observes that $z = r(x)y'(x)$ is a solution of the differential equation

$$(p^{-1}z')' + r^{-1}z = 0, \tag{5}$$

which is of the form (1).

*Example:* To study the differential equation (3)′ $y'' + a^2x^{-2}y = 0$ $(4k^2 = 4a^2 - 1 > 0)$ apply Corollary 2 with $\alpha = {}^1/_2$. Thus $|y(x)| \leq Mx^{1/2}$. Equation (5) becomes $(x^2a^{-2}z')' + z = 0$, which under the transformation $z = ax^{-1}w$ becomes $w'' + a^2x^{-2}w = 0$. Thus $|w| \leq Mx^{1/2}$ and $|y'(x)| \leq Mx^{-1/2}$. Solutions of (3)′ are readily seen to be $x^{1/2} \sin k \log x$ and $x^{1/2} \cos k \log x$.

The bibliography of this problem is voluminous. Further references to the literature will be found in the list of papers which follows.

Bellman, Richard, "The Boundedness of Solutions of Linear Differential Equations," *Duke Math. J.*, **14**, 83–97 (1947).

Boas, M., Boas, R. P., Jr., and Levinson, N., "The Growth of Solutions of a Differential Equation," *Ibid.*, **9**, 847–853 (1942).

Hille, Einar, "Non-Oscillation Theorems," *Trans. Am. Math. Soc.*, **64**, 234–252 (1948).

Levinson, Norman, "The Asymptotic Behavior of a System of Linear Differential Equations," *Am. J. Math.*, **68**, 1–6 (1946).

Weyl, H., "Comment on the Preceding Paper," *Ibid.*, **68**, 7–12 (1946).

Wintner, A., "The Adiabatic Linear Oscillator," *Ibid.*, **68**, 385–397 (1946).

# PRINCIPAL QUADRATIC FUNCTIONALS AND SELF-ADJOINT SECOND-ORDER DIFFERENTIAL EQUATIONS

BY WALTER LEIGHTON

DEPARTMENT OF MATHEMATICS, WASHINGTON UNIVERSITY

Communicated by Marston Morse, February 15, 1949

1. *Principal Quadratic Functionals.*—A *principal* quadratic functional is defined to be a functional of the form

$$J = \int_0^b [r(x)y'^2 - p(x)y^2]dx \qquad (b > 0), \tag{1}$$

where $r(x)$ and $p(x)$ are continuous and $r(x) > 0$ on the interval $0 < x \leq b$. A function $y(x)$ and the curve $y = y(x)$ are said to be $A$-admissible[1] on $0 \leq x \leq b$ if

1. $y(x)$ is continuous on $0 \leq x \leq b$ and $y(0) = y(b) = 0$;
2. $y(x)$ is absolutely continuous and $y'^2(x)$ is in $L$ on each closed subinterval of $0 < x \leq b$.

THEOREM 1. *If $p(x) > 0$ when $x$ is near zero and if*

$$\lim_{x \to 0^+} \int_x^b p(x)dx = +\infty,$$

*there exists an $A$-admissible curve $y = y(x)$ along which $J$ assumes a finite negative value.*

The above theorem generalizes certain earlier results.[1]

THEOREM 2. *If $p(x) > 0$ near $x = 0$, if there is no point on $0 \leq x < b$ conjugate to $x = 0$, and if*

$$\lim_{x \to 0^+} \int_x^b p(x)dx$$

*is finite, then*

$$\liminf_{x \to 0^+} \int_x^b (ry'^2 - py^2)dx \geq 0 \tag{2}$$

*for all $A$-admissible functions $y(x)$.*

Inasmuch as $x = 0$ is, in general, a singular point of $J$, the classical definition of the conjugate point is not valid. The definition used here is the generalization of the usual concept which was employed in a paper by Morse and Leighton.[1]

Let $y(x) \not\equiv 0$ be an extremal orthogonal to the line $x = x_0 > 0$. If $x = a$ is the first zero following $x = x_0$, lim $a$ as $x_0 \to 0+$, when it exists, is called the *focal point of the y-axis.* It can be shown that *if the y-axis contains its focal point, there exists an $A$-admissible curve $y = y(x)$ along which $J$ assumes a finite negative value.*

The *focal point solution* $y(x)$ of the Euler equation of $J$ may be defined (when $x = 0$ is not its own first conjugate point). Its smallest positive zero, if it exists, is the focal point of the $y$-axis, and conversely.

A curve $y = y(x)$ is said to be $F$-admissible[1] if

1. $y(x)$ is continuous on $(0, b]$ and $y(b) = 0$;
2. $y(x)$ is absolutely continuous and $y'^2(x)$ is in $L$ on each closed subinterval of $(0, b]$.

Theorem 3. *If* $p(x) > 0$ *near* $x = 0$, *the limit* (2) *is* $\geq 0$ *for all* $F$-*admissible curves* $y = y(x)$ *if and only if there is no focal point of the* $y$-*axis on* $0 \leq x < b$.

2. *The Oscillation of Solutions of a Self-Adjoint Second-Order Linear Differential Equation; a New Normal Equation.*—We study the detection of an infinity of oscillations of the solutions of the equation

$$[r(x)y']' + p(x)y = 0, \tag{3}$$

where $r(x)$, $r'(x)$ and $p(x)$ are continuous and $r(x) > 0$ on $0 < x < \infty$.

The following theorem appears to be fundamental.

Theorem 4. *If* $p(x) > 0$ *near* $x = +\infty$, *and if*

$$\int_1^{\infty} \frac{dx}{r(x)} = \int_1^{\infty} p(x)dx = +\infty,$$

*every solution of* (3) *has an infinity of zeros on the interval* $1 < x < \infty$. *Similarly, if* $p(x) > 0$ *near* $x = 0+$, *and if*

$$\int_0^1 \frac{dx}{r(x)} = \int_0^1 p(x)dx = +\ \infty$$

*every solution of* (3) *has an infinity of zeros on the interval* $0 < x < 1$.

The proof is elementary.

Roughly speaking the theorem is most sensitive when $\int dx/r(x)$ diverges slowly. This suggests the study of the *normal equation*

$$(xy')' + p(x)y = 0 \tag{4}$$

as superior for many purposes to that of the classical form $y'' + p(x)y = 0$. Equation (3) can be transformed into (4) by the substitution $y = (x/r)^{1/2}z$.

The following result is valid.

Theorem 5. *If* $p(x) > 0$ *and* $O(x^{-\alpha})(\alpha > 1)$ *for* $x$ *large, a non-null solution of* (4) *has at most a finite number of zeros on* $1 < x < \infty$. *Similarly, if* $p(x) > 0$ *and* $O(x^{-\alpha})(\alpha < 1)$ *near* $x = 0+$, *a non-null solution of* (4) *has at most a finite number of zeros on* $0 < x < 1$.

Detailed proofs of the theorems stated and of others, as well as a more complete bibliography, will be published elsewhere.

---

[1] Morse, M., and Leighton, W., "Singular Quadratic Functionals," *Trans. Am. Math. Soc.*, **40**, 252–286 (1936).

# *THE MECHANISM OF PROTEINURIA**

BY T. ADDIS

STANFORD UNIVERSITY MEDICAL SCHOOL, SAN FRANCISCO†

Read before the Academy, November 16, 1948

During the last war the Office of Scientific Research and Development gave this laboratory an opportunity to observe some effects on the kidney of various materials that had been suggested as possible substitutes for blood plasma lost in battle casualties. One of the methods we used was to inject large amounts of the proposed blood substitute into the peritoneal cavities of 150-g. ♀ rats (960 mg. of the material in 6 per cent solution). When a single injection was given we found, in consonance with the work of others,[1] that the rate at which the kidneys excreted any protein depended, first of all, on its molecular weight. Thus proteins whose molecular weight was about 34,000 (egg albumin, globin and Bence Jones protein) were rapidly excreted, proteins with a molecular weight of 70,000 (preparations of bovine or human serum albumin) gave only a slight and delayed increase over the usual amount of protein in the rat's urine, while globulins (molecular weight about 150,000) were not excreted at all. However, when we began to give repeated injections, we found that even proteins with a molecular weight of 70,000 began to appear in the urine until the rate of excretion rose to very high levels and the kidneys were excreting in 24 hours more than their own weight of protein.

Clinicians are called on to treat patients who excrete considerable amounts of protein in their urine. The current supposition is that in these people disease of one sort or another has damaged the glomerular membrane so that it has become permeable to protein. The absence of protein in the urine of healthy people is seen as a consequence of the fact that the normal membrane allows no protein to pass into the urine. This supposition is based on the classical experimental work of Wearn and Richards[2] who first demonstrated that the glomerulus is a filter when they succeeded in obtaining minute amounts of fluid from Bowman's capsule and showed that it was a filtrate of the plasma perfusing the glomerulus, since the concentration of various small molecule solutes was the same in the fluid as in the plasma. This filtrate they found to be "protein-free," or, at least, it contained less than 30 mg. of protein per 100 cc., which was the lowest concentration that could be detected with the most sensitive tests they used. There is thus a sound foundation for the present clinical idea about proteinuria. It might even be regarded as beyond assail if it were not that the volume filtered by a man in 24 hours is about 180,000 cc., so that even if the concentration of protein were only 10 mg. per 100 cc., i.e., too small to be detected, 18,000 mg. would every day leak through the membrane.

We agree that this simple hypothesis as to the mechanism of proteinuria should not be discarded simply because human or bovine proteins of 70,000 molecular weight, do, in the end, and after repeated injections, induce intense proteinuria in the rat. These are "foreign" proteins. We cannot be sure that they may not appear in the urine because they somehow render the glomerular membrane more permeable to protein. It is another matter when we find that proteins from the rat's own blood serum induce the excretion of large quantities of protein, particularly when the injections of serum appear to leave the animals wholly undisturbed. For in this case it would appear that the injections have done no more than increase the concentration of the proteins that are always flowing through the glomerular capillaries. These considerations led us recently to repeat and extend our first observations on the effect of injecting rat serum into rats that we made some years ago during the war.

TABLE 1

PROTEINURIA FOLLOWING EACH INJECTION OF 5 ML. OF RAT SERUM PER 100 SQ. CM OF BODY SURFACE (6.0 G. PROTEIN PER 100 CC.) EVERY 48 HRS. CONTROLS GIVEN THE SAME VOLUME OF 0.85 PER CENT NACL SOLUTION

| DAY | INJECTION | URINE PROTEIN | | SERUM PROTEIN | |
|---|---|---|---|---|---|
| | | RAT SERUM, MG. PER 24 HRS. | NACL, MG. PER 24 HRS. | RAT SERUM, G. PER 100 CC. | NACL, G. PER 100 CC. |
| 1st | + | 7 | 2 | 7.2 | 5.7 |
| 2nd | .. | 3 | 3 | ... | ... |
| 3rd | + | 19 | 3 | ... | ... |
| 4th | .. | 3 | 2 | ... | ... |
| 5th | + | 18 | 2 | 8.3 | 6.1 |
| 6th | .. | 7 | 2 | ... | ... |
| 7th | + | 23 | 2 | ... | ... |
| 8th | .. | 5 | 2 | ... | ... |
| 9th | + | 45 | 2 | 8.5 | 5.7 |

In table 1 we give average 24-hour rates of excretion per rat obtained from two groups, each containing 6 ♀ rats of about 150 g. body weight. The protein was determined by a modified biuret method.[3] One group was given intraperitoneal injections of rat serum and the other, the control group, injections of the same volume of 0.85 per cent NaCl solution.

To us the significance of the data given in table 1 lies in the fact that the slight but very definite and increasing excretion of protein that followed each successive injection of rat serum was found in a group of rats that seemed to be in better health than the control rats given salt solution. At the end of the ten-day period the average control body had fallen from 149 to 138 g. while the rat serum injected group that started at 150 g. had a body weight of 148 g. Both groups were on a protein-free diet of 35 per cent glucose in 0.4 per cent NaCl with 1 per cent of a mixture of B vitamins. Since we have found that the intraperitoneal injection of horse serum into rats prevents the loss of protein that otherwise occurs on a glucose diet[4] we suppose that the condition of the serum injected rats was better because they were able to use for anabolism some of the rat serum protein we injected. We stress this point because if the injections had disturbed the health of the animals the proteinuria might have been a consequence of damage to the glomerular membrane. It is true, of course, that the injections increased the concentration of protein in the blood. The hematocrit changes indicated that one of the effects was a dilution of the blood in respect of its red blood cell content. But these changes are neither sudden nor extreme, because the absorption of the serum from the peritoneal cavity occurs slowly over a period of several hours. Neither the rate of urine flow nor the blood serum urea concentrations indicated any deviation from normal renal function. This was a pure proteinuria. Under these circumstances we think it would be far-fetched to postulate a change in glomerular permeability due to injury. We think that some protein was always escaping through the membrane and that the rate of escape was increased by the increase in serum protein concentration that we induced.

The intensity of the proteinuria induced by rat serum injections is greatly increased if the injections are given at shorter time intervals. When 4 injections of the same volume of serum containing 6.2 g. per 100 cc. are given over a period of 48 hours steeply rising rates of excretion are observed, until after the fourth injection rates corresponding to over 400 mg. of protein per 24 hours are delivered from kidneys that contain only 160 mg. of protein—a rate, in respect of the size of the kidney, higher than we ever obtain from human kidneys no matter to what degree disease may have damaged the glomerular membrane. These very high rates per gram of kidney are associated with unusually high levels of serum protein concentration, sometimes exceeding 12 g. per 100 cc. Here, again, we find no evidence of injury to the kidney, either in the formed elements in the urine or in the histological appearance of the glomeruli. There is, however, an increase in the weight of the kidney, which is due, predominantly, to an increase, not of water, but of protein. Even with lesser degrees of proteinuria this increase in the weight of the kidneys occurs if the injections are reiterated. Thus in the case of the rats whose protein excretion after 5 injections is given in table 1, the salt injected group had kidneys that weighed 833 mg., while the group given rat serum had kidneys that weighed 1042 mg. We believe this weight increase is a consequence of the reabsorption of filtered protein by the proximal tubules.

The rat is not the only animal in which proteinuria can be induced by injection of its own plasma proteins. Thorn and his co-workers found that patients with liver cirrhosis excreted considerable quantities of protein after reiterated intravenous injections of human serum albumin[5] and recently Whipple and his associates observed in dogs injected with dog plasma long-continued and high urinary excretion rates of albumins and globulins, with no evidence of glomerular damage.[6] Admittedly such observations scarcely warrant positive conclusions. They are too indirect. Their value is negative and lies in the fact that they permit us to doubt the direct and seemingly compelling evidence given by Wearn and Richards that the glomerular filtrate is protein free and the normal glomerular membrane impermeable to serum albumin or any larger plasma protein. Nevertheless that freedom is important because once the idea is entertained that under normal conditions some protein may always be escaping into the urine we find many well-documented clinical, experimental and pathological observations can then for the first time be viewed as consequences of a quantitative intensification of processes that are physiological. We refer to the data on proximal tubule reabsorption of urine protein given by Gérard,[7] Oliver,[8] Dock,[9] Randerath,[10] Smetana,[11] and Yuile and others[12] and to the work of Rather,[13] on the digestion of hemoglobin reabsorbed from the urine. The clinician who has to try to comprehend what he observes in patients with proteinuria finds that there are many questions that have to be laid aside as insoluble so long as he holds to the postulate of the absolute impenetrability to protein of the normal glomerular membrane (capillary wall, basement membrane and epithelial membrane), and continues to believe that all proteinurias arise only because of damage to the membrane. It remains true, of course, that the more complex idea of the mechanism of proteinuria as involving three processes, the filtration, the reabsorption and the intracellular digestion of protein, remains no more than a working hypothesis. But the advantage of the more complex view is that it is susceptible of experimental validation, correction and extension, though that will require the coördinated efforts of groups of workers employing the techniques of physical chemistry, histochemistry and enzyme chemistry, in conjunction with the simpler methods of experimental biology.

---

* This work was made possible by a grant from the Public Health Service.

† The author has removed to Institute for Medical Research, Cedars of Lebanon Hospital, 4751 Fountain Ave., Los Angeles 27, California.

[1] Bayliss, L. E., Kerridge, P. M. T., and Russell, D. S., *J. Physiol.*, **77**, 386 (1933). Bott, P. A., and Richards, A. N., *J. Biol. Chem.*, **141**, 291 (1941).

[2] Wearn, J. T., and Richards, A. N., *Am. J. Physiol.*, **71**, 209 (1924).

[3] Addis, T., Barrett, E., Boyd, R. I., and Ureen, H. J., *J. Exp. Med.*, **89**, 131 (1949).

[4] Addis, T., Lee, D. D., Lew, W., and Poo, L. J., *Am. J. Physiol.*, **128**, 544 (1939–1940).

[5] Thorn, G. W., Armstrong, S. H., Jr., and Davenport, V. D., *J. Clin. Invest.*, **25**, 304 (1946).

[6] Terry, R., Hawkins, D. R., Church, E. H., and Whipple, G. H., *J. Exp. Med.*, **87**, 561 (1948).

[7] Gérard, P., *J. Anat.*, **70**, 354 (1935).

[8] Oliver, J., *Harvey Lectures, Series XL*, 1944–1945, p. 102. *J. Mount Sinai Hosp.* **15**, 175, (1948).

[9] Dock, W., *New Eng. J. Med.*, **227**, 633 (1942).

[10] Randerath, E., *Ergeb. d. Allg. Path. u. Path. Anat. des Menschen u. der Tiere*, **32**, 91 (1937).

[11] Smetana, H., *Am. J. Path.*, **18**, 1029 (1942) and **23**, 255 (1947).

[12] Monke, J. V., and Yuile, C. L., *J. Exp. Med.*, **72**, 149 (1940). Yuile, C. L., and Clark, W. F., *Ibid.*, **74**, 187 (1941). Yuile, C. L., Steinman, J. F., Hahn, P. F., and Clark, W. F., *Ibid.*, **74**, 197 (1941).

[13] Rather, L. J., *Ibid.*, **87**, 163 (1948).

---

# *CRYSTALLOGRAPHIC SYMMETRY IN RECIPROCAL SPACE**

BY M. J. BUERGER

CRYSTALLOGRAPHIC LABORATORY, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, DEPARTMENT OF GEOLOGY

Communicated by W. J. Mead, February 11, 1949

Reciprocal space constitutes a useful tool for dealing with many diffraction problems,[1] particularly problems concerned with the diffraction of x-rays by crystals. Wrinch[2] has shown that if a distribution has a point-group symmetry element through the origin, its Fourier transform in reciprocal space has the same symmetry element. For crystallographic purposes, it is desirable to go farther than this and inquire how the more general crystallographic symmetry elements appear in reciprocal space.

Consider a symmetry element with a translation component, such as $A$, an $n$-fold screw through the origin, $O$, figure 1. Let the screw, for definiteness, lie along the crystallographic $c$ axis. The figure shows two sets of planes related by an operation of the screw. One set of planes may be indexed $hkl$, the other $h'k'l'$, but both have the common interplanar spacing $d_{hkl}$. In reciprocal space, these sets of planes are represented by points $P$ and $P'$, respectively. $P$ lies on the normal to $hkl$, and $P'$ lies on the normal to $h'k'l'$, both normals extending from the origin. Both $P$ and $P'$ have the same origin distance, namely $1/d_{hkl}$. Since the direction of the normal is unaffected by the displacement of the planes due to the translational component of the axis, it is evident that the rotational component of the crystal symmetry element is preserved in the relative locations of the points $hkl$ and $h'k'l'$ in reciprocal space.

The relative phases of the two points, however, depend on the relative displacements of the two sets of planes from the origin.[1] In figure 1, the phase difference in reciprocal space corresponding with the displacement between two neighboring planes of the set $hkl$ is $2\pi$. This difference can be measured by the displacement of the planes along the $c$ axis. Since there are $l$ planes per unit translation along the $c$ axis, the total phase difference per unit translation along the $c$ axis is $2\pi l$. However, the translation component of the screw produces a displacement of planes only a fraction of a translation, namely $(1/q)c$, so that the phase difference due to this translation component is $2\pi l/q$.

Let the reciprocal at $P$ be $F_{hkl}$. In general, the phase of $F_{hkl}$ is not zero, but some value, say $\varphi$. Thus the phase component of $F_{hkl}$ in the complex plane is $e^{2\pi i\varphi}$. Then the phase component at $P'$ according to the last paragraph is $e^{2\pi i\varphi}e^{2\pi il/q}$. Since the absolute magnitudes at $P$ and $P'$ are identical, the reciprocals at these two points must be related as follows:

$$F_{h'k'l'} = F_{hkl}e^{2\pi il/q},$$

where $c/q$ is the translation component of the symmetry operation. Note that the phase difference vanishes for symmetry elements without a translation component (i.e., $q = 0$), and also for indices such that $l/q$ = an integer.

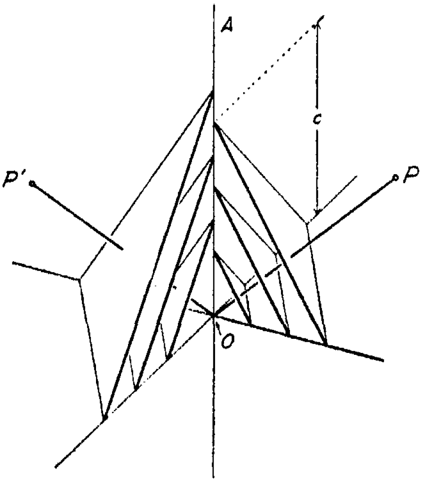


FIGURE 1

By permitting the rotational component of the symmetry element to include improper rotations, this result can be extended to include symmetry operations of the second sort. Alternatively, corresponding results can be derived directly for reflections, glide-reflections and roto-inversions. Neither the sequence of planes from the origin nor their distances away from the origin are affected by the improper character of the rotation, so that the only effect that the nature of the rotation has on reciprocal space is with respect to the location of the point.

As a consequence of these considerations, the following statement can be made:

Theorem 1. *If a crystal contains a symmetry element through a chosen* *origin in direct space, its reciprocal has the isomorphous (translation-free) symmetry element at the origin, but the points in the fields related by the operations of the symmetry element have phases related by* $e^{2\pi i p h/q}$, *where* $p$ *is the power of the operation relating the fields and* $a/q$ *is the translation component of the operation.*

(For diagonal symmetry elements having the general translation component $(a + b + c)/q$, it can be shown that $(h + k + l)$ should be substituted for $h$.)

In case the symmetry axis does not contain the origin, but is displaced to a location having coördinates $xa$, $yb$, the planes $hkl$ (along with their symmetrically equivalent planes) are displaced too, as shown in figure 2. Since there are $h$ planes along $a$, and each separation corresponds with a phase difference in reciprocal space of $2\pi$, a displacement of the entire $a$ translation corresponds to $2\pi h$. Therefore, the component of the displacement of the axis along $a$, of amount $xa$, corresponds with a phase difference of $2\pi xh$. Similarly, the component of its displacement, of amount $yb$, along $b$ corresponds with a phase difference of $2\pi yk$. Consequently, the entire displacement of the axis from the origin to location $xa$, $yb$ corresponds with a phase difference of $2\pi(xh + yk)$ applied to $F_{hkl}$ in reciprocal space. Symbolically,

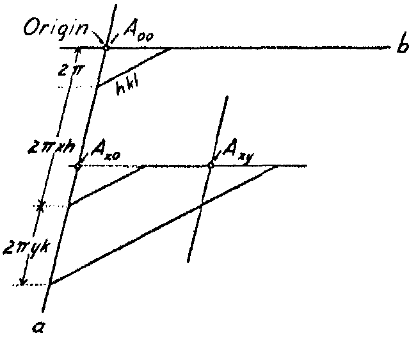


FIGURE 2

$${}_{x,y}F_{hkl} = {}_{0,0}F_{hkl}e^{2\pi i(xh + yk)}.$$

THEOREM 2. *If a symmetry element is displaced by an amount* $xa + yb$, *the phase component difference produced on* $F_{hkl}$ *accompanying this displacement is* $e^{2\pi i(xh + yk)}$.

Space groups can be developed by adding to the lattice operations various generating operations isomorphous with the point-group operations. In the reciprocal of the space group, the translations of the lattice are responsible for limiting the non-vanishing values of the Fourier transform of the crystal structure to the points of the reciprocal lattice in reciprocal space. The generating symmetry operations and their locations are responsible for fixing the relative phases at the points of the reciprocal lattice which are equivalent by the point-group symmetry.

THEOREM 3. *For each generating operation of the cell of the space group, there is a symmetry relation between points in reciprocal space: The position and absolute-magnitude symmetry is the same as that of the point group isomorphous with the space group, non-vanishing values occurring only at points of the lattice reciprocal to the lattice of the crystal, while the phases of the points related by this symmetry are governed by the translation components and positions of the symmetry elements as well as the indices of the points (in accordance with Theorems 1 and 2).*

These theorems make it possible to write down by inspection, all the relations between $F$'s for a given space group for any chosen origin. The customary method of obtaining such information is the laborious structure-factor method.[3] Theorem 2 also provides a tool for predicting the changes in phase relations which accompany any displacement of the origin of the space group.

A space group is determined by the lattice type, the point group and the generating operations isomorphous with the point-group operations. Since each of these has a unique representation in reciprocal space for a chosen origin, it is evident that all space groups can be distinguished by their representations in reciprocal space.

---

[*] This paper contains the essence of a lecture given at the Faculdade Nacional de Filosofia, Rio de Janeiro, December 24, 1948, on the subject, "Reciprocal Space."

[1] Garrido, Julio, "El Espacio Recíproco y su Aplicacíon a Algunos Problemas de la Difusión de los Rayos X," *Real Academia de Ciencias de Madrid*, Tomo II, Memoria No. 1, 1947.

[2] Wrinch, Dorothy, "Fourier Transforms and Structure Factors," *A.S.X.R.E.D Monograph*, No. 2, February, 1946, pp. 10–12.

[3] Lonsdale, Kathleen, *Simplified Structure Factor and Electron Density Formulae for the 230 Space Groups of Mathematical Crystallography*, G. Bell & Sons, Ltd., London, 1936, pp. 7-12.

---

# *PHYSICAL CURVES IN SPACE OF n DIMENSIONS*[*]

BY EDWARD KASNER AND JOHN DE CICCO

DEPARTMENTS OF MATHEMATICS, COLUMBIA UNIVERSITY, NEW YORK, AND ILLINOIS INSTITUTE OF TECHNOLOGY, CHICAGO

Communicated February 25, 1949

1. A *system* $S_k$ of $\infty^{2n-1}$ curves in a given positional field of force[1] in Euclidean space of $n$ dimensions consists of curves along which a constrained motion is possible such that the osculating plane at each point contains the force vector $F$, and the pressure $P$ (along the principal normal

to the curve) is proportional to the normal component $N$ of $F$. Thus $P = kN$ where $k(\neq -1)$ is the constant of proportionality.[2]

*The vector differential equation of a system* $S_k$ *of* $\infty^{2n-1}$ *curves is*

$$\ddot{x} = (1 + k)F - \frac{k(F\dot{x})}{\dot{x}^2}\,\dot{x}. \tag{1}$$

2. Let $\lambda_1, \lambda_2, \ldots, \lambda_n$, be the mutually orthogonal set of principal unit vectors, such that the scalar product $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is unity, at a point $x = x^\alpha (\alpha = 1, \ldots, n)$ of a curve $C$. The Serret-Frenet formulas are

$$\frac{d\lambda_\beta}{ds} = -\kappa_{\beta-1}\lambda_{\beta-1} + \kappa_\beta\lambda_{\beta+1}, \tag{2}$$

for $\beta = 1, \ldots, n$. Of course, $\kappa_0 = 0$ and $\kappa_n = 0$. The quantities $\kappa_1, \kappa_2, \ldots, \kappa_{n-1}$ are the $(n - 1)$ curvatures of $C$ at $x$.

By means of (2), it is found that *a system of* $S_k$ *of* $\infty^{2n-1}$ *curves, as defined by (1), is given by the equations*

$$\dot{v} = T, \qquad v^2\kappa_1 = (1 + k)N, \qquad F = T\lambda_1 + N\lambda_2, \tag{3}$$

*where* $T = (F\lambda_1)$ *and* $N = (F\lambda_2)$.

Eliminating the speed $v$ from (3), we find that *the intrinsic equations of a system* $S_k$ *of* $\infty^{2n-1}$ *curves are*

$$N\frac{d\kappa_1}{ds} = \left(\frac{dF}{ds}\lambda_2\right)\kappa_1 - \frac{3 + k}{1 + k}T\kappa_1^2, \qquad F = T\lambda_1 + N\lambda_2. \tag{4}$$

*The rate of variation* $d\rho_1/ds$ *of the first radius of curvature* $\rho_1 = 1/\kappa_1$ *per unit length of arc is a linear integral function of* $\rho_1$ *with coefficients depending on* $x$, $\lambda_1$, $\lambda_2$ *only. That is,*

$$N\frac{d\rho_1}{ds} = -\left(\frac{dF}{ds}\lambda_2\right)\rho_1 + \frac{3 + k}{1 + k}T. \tag{5}$$

3. Consider any curve $C$ along which a constrained motion is possible such that the osculating plane at each point of $C$ contains the force vector $F$. We denote the totality of all such curves $C$ by $\Sigma$. Thus $S_k$ is a subset of $\Sigma$.

*At each point of a curve* $C$ *of* $\Sigma$, *the osculating m flat where* $m < n$ *contains the vector* $d^{m-2}F/ds^{m-2}$.

All the curves $C$ of $\Sigma$ which pass through a given point in a given direction have the same osculating plane and the same osculating three flat at this point provided that $N \neq 0$. The osculating $m$ flat where $m = 4, \ldots, n - 1$, varies.

*Let* $N \neq 0$. *The second curvature* $\kappa_2$ *is a function of* $x$, $\lambda_1$, $\lambda_2$ *only. If*

*$\kappa_2 \neq 0$, the third curvature $\kappa_3$ is the square root of a quadratic polynomial in $\kappa_1$ with coefficients functions of $x$, $\lambda_1$, $\lambda_2$ only.*

4. Consider the curvatures $\kappa_1, \kappa_2, \ldots, \kappa_{n-1}$ of the curves of a system $S_k$. As $S_k$ is a special case of $\Sigma$, it is seen that $\kappa_2$ and $\kappa_3$ have the properties described above.

In general, *the m-th curvature $\kappa_m$ for $m = 3, 4, \ldots, n - 1$, of any curve C of a system $S_k$, is the square root of a polynomial in $\kappa_1$ divided by a polynomial in $\kappa_1$.*

The initial curvatures of the $\infty^1$ curves of a system $S_k$ which pass through a fixed point in a fixed direction obey the following laws. The first curvature $\kappa_1$ varies inversely as the square of the speed $v$. The second curvature $\kappa_2$ is constant. The third curvature $\kappa_3$ is the square root of a quadratic polynomial of $\kappa_1$. The fourth curvature $\kappa_4$ is the square root of a fourth degree polynomial of $\kappa_1$ divided by the quadratic polynomial appearing in $\kappa_3$. The fifth curvature $\kappa_5$ is the square root of a tenth degree polynomial of $\kappa_1$ divided by the fourth degree polynomial appearing in $\kappa_4$.

5. If a particle starts from rest, the resulting curve of the system $S_k$ is called a *rest trajectory*. The vector $d^{m-1}F/ds^{m-1}$ is in the osculating $m$ flat for $m = 1, \ldots, n - 1$, at the initial point of a rest trajectory.

*The rest trajectory and the line of force are initially tangent and possess the same osculating m flat for $m = 2, 3, \ldots, n - 1$.*

The vector $dF/ds$ for the rest trajectory coincides with the vector $dF/ds$ for the line of force. However, for $m = 3, 4, \ldots$, the vector $d^{m-1}F/ds^{m-1}$ constructed for the rest trajectory is distinct from the corresponding vector $d^{m-1}F/ds^{m-1}$ constructed for the line of force. These two vectors are situated in the same $m$ flat.

*The ratio of the first curvature of the rest trajectory to the first curvature of the line of force is*

$$\frac{1 + k}{3 + k}. \qquad (6)$$

For dynamical trajectories ($k = 0$), this ratio is 1/3, a well-known result. For general catenaries ($k = 1$), this ratio is 1/2. For brachistochrones ($k = -2$), this ratio is $-1$. For velocity curves ($k = \infty$), this ratio is $+1$.

There is no analogous such formula for the second curvatures or torsions, or for the remaining curvatures.[3]

6. The center $S$ of the osculating $m$ sphere of a curve $C$ is

$$S = x + \rho_1\lambda_2 + \rho_2\frac{d\rho_1}{ds}\lambda_3 + \rho_3\left[\frac{\rho_1}{\rho_2} + \frac{d}{ds}\left(\rho_2\frac{d\rho_1}{ds}\right)\right]\lambda_4 + \ldots + \sigma_{m+1}\lambda_{m+1}, \qquad (7)$$

where $\rho_1, \ldots, \rho_{n-1}$ are the $(n - 1)$ radii of curvature and the various coefficients may be obtained by the recursion formula

$$\sigma_p = \rho_{p-1}\left[\frac{\sigma_{p-2}}{\rho_{p-2}} + \frac{d\sigma_{p-1}}{ds}\right] \qquad (8)$$

for $p = 2, 3, \ldots, m+1$.

*The locus of the centers of the osculating m spheres of the* $\infty^1$ *curves of a system* $S_k$ *which pass through a fixed point in a fixed direction is an algebraic curve whose position vector S is an algebraic function of the first curvature* $\kappa_1 = 1/\rho_1$.

*For* $m = 2$*, this locus is a straight line, just as is the case for* $n = 3$. This line is given by the system of $(n-1)$ equations

$$\lambda_1(S - x) = 0, \quad \lambda_4(S - x) = 0, \quad \ldots, \quad \lambda_n(S - x) = 0,$$
$$(1 + k)\frac{dF}{ds}(S - x) = (3 + k)(F\lambda_1). \qquad (9)$$

To each direction $\lambda_1$ through the fixed point $x$, there corresponds a straight line (9). The line which corresponds to the direction of the force vector acting at the point $x$ is parallel to the binormal of the line of force. Its distance from the point $x$ is $(3 + k)/(1 + k)$ times the radius of curvature of the line of force.

This distance from the fixed point $x$ is equal to the radius of curvature of the rest trajectory through the fixed point. Hence the above line passes through the center of circular curvature of the rest trajectory, a new general theorem for all physical families.

* Presented to the American Mathematical Society, 1949.

[1] Kasner, *Differential Geometric Aspects of Dynamics*, Amer. Math. Soc. Publications 1913, 1934, 1948. See the series of papers in *Trans. Amer. Math. Soc.*, **7**–11 (1906–1910).

[2] Kasner, "Physical Curves," PROC. NAT. ACAD. SCI., **33**, 246–251 (1947).

[3] Kasner and Mittleman, "A General Theorem on the Initial Curvatures of Dynamical Trajectories," *Ibid.*, **28**, 48–52 (1942). De Cicco, "Extensions of Certain Dynamical Theorems of Halphen and Kasner," *Bull. Amer. Math. Soc.*, **49**, 736–744 (1943); and "Constrained Motion upon a Surface Under a Generalized Field of Force," *Ibid.*, **53**, 993–1001 (1947). Additional theorems and relations to the problem of three bodies will be given in a future paper.

---

# ON BENDING OF CURVED THIN-WALLED TUBES*

BY ERIC REISSNER

DEPARTMENT OF MATHEMATICS MASSACHUSETTS OF TECHNOLOGY

Communicated by T. von Kármán, February 11, 1949

1. *Introduction.*—The present note is concerned with the problem of pure bending of curved tubes. The fact that a curved thin-walled tube is more flexible than a straight tube of the same cross-section has first been explained by von Kármán[1] who showed that the flattening of the

tube cross-sections due to the action of the primary bending stresses is responsible for this effect. Von Kármán's approximate solution of the problem makes use of the minimum principle for the displacements in elasticity in Rayleigh-Ritz fashion. Subsequently H. Lorenz[2] gave an alternate approximate solution by the use of the minimum principle for the *stresses* in elasticity. This same method was also used by L. Beskin[3] in a recent paper on the subject.

In what follows we propose to consider the problem from the point of view of the theory of thin elastic shells. Inasmuch as the bending of a curved thin tube may be thought of as a problem of determining an axisymmetrical stress distribution in a shell of revolution it must be possible to relate the resultant formulation of the problem to the results of H. Reissner[4] and E. Meissner[5] on the bending and stretching of shells of revolution. H. Reissner, E. Meissner and their followers restricted attention to problems in which not only stresses but also displacements are rotationally symmetric and univalued. Evidently the latter assumption excludes the problem of tube bending from their results. In what follows we make the necessary slight generalization of the theory such that a system of equations is obtained which contains new terms originating from a suitably chosen multivalued expression for one of the components of displacement. We may remark that in so doing we prepare the way for treating the problem of pure tube bending in a manner which is analogous to M. Golovin's treatment[6] of pure bending of a circular ring sector of uniform thickness according to the theory of plane stress.

*2. Basic Relations for Axisymmetrical Deformations of Thin Shells of Revolutions.*—Let $r = r(\xi)$, $z = z(\xi)$ be the equations of the middle surface of the shell in parametric form and write

$$\alpha^2 = (r')^2 + (z')^2, \qquad r' = \alpha \cos \phi, \qquad z' = \alpha \sin \phi, \tag{1}$$

where primes indicate differentiation with respect to $\xi$ and where $\phi$ is the tangent angle to a meridian curve. Let $u$ and $w$ be components of displacement in *radial* and *axial* direction, respectively, and let $\beta$ be the change of sloping angle due to deformation. It can then be shown that for small deformations and under the assumption that normals to the middle surface are preserved during deformation the state of strain in the deformed shell may be described as follows:

$$\epsilon_\xi = \epsilon_{\xi M} + \zeta\kappa_\xi, \qquad \epsilon_\theta = \epsilon_{\theta M} + \zeta\kappa_\theta, \tag{2}$$

where $\zeta$ is the distance from the middle surface and where

$$\left.\begin{aligned} \epsilon_{\xi M} &= \frac{u'}{r'} - \beta \tan \phi, \qquad \kappa_\xi = \frac{\beta'}{\alpha} \\ \epsilon_{\theta M} &= \frac{u}{r}, \qquad \kappa_\theta = \frac{r'}{r}\,\frac{\beta}{\alpha}. \end{aligned}\right\} \tag{3}$$

Equations (3) imply a compatibility relation of the form

$$(r\epsilon_{\theta M})' - r'\epsilon_{\xi M} = z'\beta. \qquad (4)$$

Let $H$, $V$, $N_\xi$, $N_\theta$ and $Q$ be stress results, let $M_\xi$ and $M_\theta$ be stress couples and let $p_H$ and $p_V$ be components of load intensity. We then have the following equilibrium equations:

$$\begin{gathered}(rV)' + r\alpha p_V = 0\\ (rH)' - \alpha N_\theta + r\alpha p_H = 0\\ (rM_\xi)' - r'M_\theta + z'(rH) - r'(rV) = 0\\ r\alpha N_\xi = r'(rH) + z'(rV)\\ r\alpha Q = -z'(rH) + r'(rV).\end{gathered} \qquad (5)$$

Equations (5) together with appropriate stress-strain relations are a form of the basic equations from which the results of H. Reissner and E. Meissner may be deduced.[7] Before indicating the way in which this is done we shall introduce into (3) to (5) the modification which is necessary to include the problem of pure tube bending.

*3. Introduction of Multivalued Circumferential Displacement Component.*—Equations (3) for the relevant components of strain were obtained under the assumption of vanishing circumferential displacement component $v$. We now assume instead that

$$v = kr\theta, \qquad (6)$$

where $k$ is a constant and $r$ and $\theta$ are polar coördinates in the $x$, $y$-plane. Note that when $v$ is given by (6) and $u = w = 0$ then the components of strain in cylindrical coördinates have the values $\epsilon_r = \epsilon_z = \gamma_{r\theta} = \gamma_{rz} = \gamma_{\theta z} = 0$ and $\epsilon_\theta = k$. This means that the effect of replacing the condition $v = 0$ by (6) leaves $\epsilon_{\xi M}$, $\kappa_\xi$ and $\kappa_\theta$ of equations (3) unchanged and makes

$$\epsilon_{\theta M} = \frac{u}{r} + k. \qquad (3')$$

In view of (3′) the compatibility equation (4) is changed to

$$(r\epsilon_{\theta M})' - r'\epsilon_{\xi M} = z'\beta + r'k. \qquad (4')$$

*4. The Differential Equations of the Bending Problem.*—The procedure from here on is the same as that for the case $k = 0$, that is, two simultaneous second-order equations are derived for the quantities $rH$ and $\beta$. In doing this we shall assume an orthotropic tube material with the following stress-strain relations:

$$\epsilon_{\xi M} = \frac{N_\xi}{C_\xi} - \frac{N_\theta}{C_{\xi\theta}}, \qquad \epsilon_{\theta M} = \frac{N_\theta}{C_\theta} - \frac{N_\xi}{C_{\xi\theta}} \qquad (7)$$

$$M_\xi = D_\xi\kappa_\xi + D_{\xi\theta}\kappa_\theta, \qquad M_\theta = D_\theta\kappa_\theta + D_{\xi\theta}\kappa_\xi. \qquad (8)$$

For the present problem we may in equations (5) set $p_N = p_H = V = 0$. We combine (7) and (4′) and use (5) to obtain the first of the two simultaneous equations in the following form:

$$(rH)'' + \frac{(r/\alpha C_\theta)'}{(r/\alpha C_\theta)}(rH)' - \left[\frac{C_\theta}{C_\xi}\left(\frac{r'}{r}\right)^2 + \frac{(r'/\alpha C_{\xi\theta})'}{(r/\alpha C_\theta)}\right](rH) = \frac{z'\beta + r'k}{(r/\alpha C_\theta)}. \tag{I}$$

Next we combine (8) and (3) and introduce the result into the third of equations (5) to obtain the second of the two simultaneous equations in the following form:

$$\beta'' + \frac{(rD_\xi/\alpha)'}{(rD_\xi/\alpha)}\beta' - \left[\frac{D_\theta}{D_\xi}\left(\frac{r'}{r}\right)^2 - \frac{(r'D_{\xi\theta}/\alpha)'}{(rD_\xi/\alpha)}\right]\beta = -\frac{z'(rH)}{(rD_\xi/\alpha)}. \tag{II}$$

Equations (I) and (II) must be integrated, subject to suitable boundary conditions for open tubes or periodicity conditions for closed tubes. A relation between $k$ and the applied bending moment $m$ is obtained, after (I) and (II) are solved, by means of the equation defining $m$,

$$m = \int rN_\theta \alpha d\xi - \int M_\theta z' d\xi. \tag{9}$$

5. *Some Special Cases.*—Equations corresponding to von Kármán's assumptions are obtained by setting

$$C_\xi = C_{\xi\theta} = \infty, \qquad D_\theta = D_{\xi\theta} = 0. \tag{10}$$

With these assumptions equation (I) may be reduced to the form

$$[(r/\alpha C_\theta)(rH)']' = z'\beta + r'k, \tag{11}$$

while equation (II) may be reduced to

$$[(rD_\xi/\alpha)\beta']' = -z'(rH). \tag{12}$$

An equation governing the bending of a circular ring sector is obtained from (I) by setting

$$r = \xi, \qquad z = 0, \qquad \alpha = 1, \tag{13}$$

and is of the form

$$(rH)'' + \frac{(r/C_\theta)'}{(r/C_\theta)}(rH)' - \left[\frac{C_\theta}{C_\xi}\frac{1}{r^2} + \frac{(1/C_{\xi\theta})'}{(r/C_\theta)}\right](rH) = \frac{C_\theta}{r}k. \tag{14}$$

When $C_\theta = C_\xi = Eh =$ const. and $C_{\xi\theta} = Eh/\nu =$ const. equation (14) leads to Golovin's solution[6] of the problem. When $C_\theta = Eh =$ const. and $C_{\xi\theta} = C_\xi = \infty$ then equation (14) leads to the hyperbolic distribution of $N_\theta$ first obtained by Winkler and Resal.[6]

* Prepared under the sponsorship of the Office of Naval Research of the Navy Department, under contract N5 ori-07834 with Massachusetts Institute of Technology.

[1] von Kármán, T., *Z. Ver. deut. Ing.*, **55**, 1889–1895 (1911).

[2] Lorenz, H., *Physik. Z.*, **13**, 768–774 (1912).

[3] Beskin, L., *J. Applied Mechanics*, **12**, A1–A7 (1945).

[4] Reissner, H., *Festschrift H. Mueller-Breslau*, pp. 181–193. G. Kroener, Leipzig, 1912.

[5] Meissner, E., *Physik. Z.*, **14**, 343 (1913).

[6] Timoshenko, S., *Theory of Elasticity*, p. 60. McGraw-Hill, New York, 1934.

[7] Reissner, E., *H. Reissner Anniversary Volume*, pp. 231–247, J. W. Edwards, Ann Arbor, Michigan, 1949.

---

# *NOTE ON THE THEORY OF LIFTING SURFACES**

BY ERIC REISSNER

DEPARTMENT OF MATHEMATICS, MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Communicated by T. von Kármán, February 11, 1949

*1. Introduction.*—In this note we consider the problem of the steady motion of a rectangular lifting surface of finite span in incompressible flow. We are concerned with two separate objects. One is an attempt to derive somewhat more clearly than in the original paper Weissinger's improvement of Prandtl's lifting-line theory.[1] The other is to generalize the one-dimensional integral equation of lifting-line theory for the spanwise variation of lift intensity to two simultaneous equations for lift intensity and moment intensity. In so doing, certain difficulties of lifting-line theory and of its improvements by Weissinger with reference to behavior at and near the tips of the wing are substantially ameliorated.

While we restrict attention here to a rectangular surface and to incompressible flow, the extension to subsonic compressible flow is a simple matter, and also the extension to tapered wings with or without sweep is feasible.

*2. The integral equation of the rectangular lifting surface.*—We consider a nearly plane wing with chord $2b$ and span $2sb$. The integral equation of the linearized theory may be written in the following dimensionless form:[2]

$$w(x, y) = \frac{1}{\pi}\int_{-1}^{1}\frac{p(\xi, y)}{x-\xi}\,d\xi + \frac{1}{2\pi}\int_{-s}^{s}\int_{-1}^{1}\frac{\partial p}{\partial \eta}\left\{\frac{1}{y-\eta} + K(x-\xi, y-\eta)\right\}d\xi d\eta. \quad (1)$$

In equation (1) $w$ is proportional to the slope $\partial h/\partial x$ of the lifting surface, and $p$ is the pressure jump across the lifting surface. Of all singular integrals the Cauchy principal value is to be taken. The function $K$

is of the form

$$K = \frac{\sqrt{(x-\xi)^2 + (y-\eta)^2} - |y-\eta|}{(x-\xi)(y-\eta)}. \qquad (2)$$

The problem of lifting surface theory consists in the solution of (1) for $p(x, y)$ in the region $(-1 \leqslant x \leqslant 1;\ -s \leqslant y \leqslant s)$ subject to the trailing edge condition that $p(1, y)$ remains finite.

3. *Lifting-Line Theory.*—Lifting-line theory may be thought of as an approximate solution of (1), in the following sense. The term $K$ in the integral equation is neglected, essentially on the basis of the fact that for sufficiently large values of the aspect ratio $s$ $K$ behaves like $1/(y-\eta)|y-\eta|$ and is therefore small compared with $1/(y-\eta)$ over most of the range of integration.

To obtain the lifting-line equation one may multiply the abbreviated equation (1) by a factor $[(1+x)/(1-x)]^{1/2}$ and integrate as follows:

$$\int_{-1}^{1} \sqrt{\frac{1+x}{1-x}}\, w dx = \frac{1}{\pi}\int_{-1}^{1} p(\xi, y) \left\{\int_{-1}^{1} \sqrt{\frac{1+x}{1-x}}\, \frac{dx}{x-\xi}\right\} d\xi$$
$$+ \frac{1}{2\pi}\int_{-1}^{1} \sqrt{\frac{1+x}{1-x}}\, dx \int_{-s}^{s} \frac{d}{d\eta}\left(\int_{-1}^{1} p d\xi\right)\frac{d\eta}{y-\eta}. \qquad (3)$$

Let

$$l(y) = \int_{-1}^{1} p(\xi, y) d\xi \qquad (4)$$

and use the formula

$$\int_{-1}^{1} \sqrt{\frac{1+x}{1-x}}\, \frac{dx}{x-\xi} = \pi. \qquad (5)$$

Equation (3) then assumes the following form:

$$\int_{-1}^{1} \sqrt{\frac{1+x}{1-x}}\, w dx = l(y) + \frac{1}{2}\int_{-s}^{s} \frac{dl}{d\eta}\, \frac{d\eta}{y-\eta}. \qquad (6)$$

One disadvantage of this theory consists, for wings with finite tip chord, in the fact that, while the lift intensity assumes the correct values zero at the tips, the pressures $p(x, \pm s)$ do not, in general, and the results of the theory will be of doubtful validity in regions adjacent to the tips, presumably up to distances inward from the tips of the order of magnitude of the tip chord.

This behavior of the lifting-line theory may be seen by a consideration of the section moment $m$, defined as

$$m(y) = \int_{-1}^{1} p(\xi, y)\xi d\xi. \qquad (7)$$

It can be shown as follows that, in general, $m(\pm s)$ does not assume the value zero which it would have in the solution of the complete equation

(1). We multiply the abbreviated equation (1) by a factor $(1 - x^2)^{1/2}$ and integrate again across the chord,

$$\int_{-1}^{1} \sqrt{1 - x^2}\, w dx = \frac{1}{\pi} \int_{-1}^{1} p(\xi, y) \left\{ \int_{-1}^{1} \frac{\sqrt{1 - x^2}}{x - \xi} dx \right\} d\xi$$

$$+ \frac{1}{2\pi} \int_{-1}^{1} \sqrt{1 - x^2}\, dx \int_{-s}^{s} \frac{dl}{d\eta} \frac{d\eta}{y - \eta}. \qquad (8)$$

With (7) and with the relation

$$\int_{-1}^{s1} \frac{\sqrt{1 - x^2}}{x - \xi} dx = -\pi\xi,$$

we may write (8) in the form

$$\int_{-1}^{1} \sqrt{1 - x^2}\, w dx = -m(y) + \frac{1}{4} \int_{-s}^{s} \frac{dl}{d\eta} \frac{d\eta}{y - \eta}. \qquad (10)$$

Combination of (10) and (6) gives the following alternate expression for $m$,

$$m = -\frac{1}{2} l + \frac{1}{2} \int_{-1}^{1} \sqrt{\frac{1 + x}{1 - x}}\, w dx - \int_{-1}^{1} \sqrt{1 - x^2}\, w dx. \qquad (11)$$

In general the second and third terms in (11) do not cancel each other and consequently we have in general $m(\pm s) \neq 0$. A known exception forms the case of constant $w(x, \pm s)$ that is the case of a wing with uncambered tip chord.

4. *Weissinger's Extension of the Lifting-Line Theory.*—This extension takes into account in an approximate manner the term with $K$ in equation (1). It may be derived in the following way. Corresponding to (6) and in the same manner, the following exact relation is first obtained

$$\int_{-1}^{1} \sqrt{\frac{1 + x}{1 - x}}\, w dx = l(y) + \frac{1}{2} \int_{-s}^{s} \frac{dl}{d\eta} \frac{d\eta}{y - \eta}$$

$$+ \frac{1}{2} \int_{-s}^{s} \int_{-1}^{1} \frac{\partial p}{\partial \eta} \left\{ \frac{1}{\pi} \int_{-1}^{1} \sqrt{\frac{1 + x}{1 - x}}\, K(x - \xi, y - \eta) dx \right\} d\xi d\eta. \qquad (12)$$

The last term on the right of (12) is to be taken into account approximately only and this is done by stipulating that in it one may set

$$p(\xi, \eta) = \frac{l(\eta)}{\pi} \sqrt{\frac{1 - \xi}{1 + \xi}}, \qquad (13)$$

which is the exact solution when $w$ is constant and the flow is two-dimensional. Taking as an abbreviation for the triple integral in (12) the symbol $I$ we have then

$$I = \frac{1}{2}\int_{-s}^{s}\frac{dl}{d\eta}\left\{\frac{1}{\pi^2}\int_{-1}^{1}\int_{-1}^{1}\sqrt{\frac{1-\xi}{1+\xi}}\sqrt{\frac{1+x}{1-x}}\,K(x-\xi,\,y-\eta)\,d\xi dx\right\}d\eta$$
$$= \frac{1}{2}\int_{-s}^{s}\frac{dl}{d\eta}\,\bar{K}(y-\eta)\,d\eta. \quad (14)$$

Equations (12) and (14) are not yet Weissinger's results but rather, in principle, an improvement of his results for the rectangular plan-form wing.

Weissinger's first result, called the $F$-method, is obtained by introducing the following approximation

$$\frac{1}{\pi}\int_{-1}^{1}\sqrt{\frac{1+x}{1-x}}\,K(x-\xi,\,y-\eta)\,dx \approx K\left(\frac{1}{2}-\xi,\,y-\eta\right). \quad (15)$$

Note that equation (15) would be exact if $K$ were a linear function of $x$. With (15) we have further

$$\bar{K}(y-\eta) \approx \frac{1}{\pi}\int_{-1}^{1}\sqrt{\frac{1-\xi}{1+\xi}}\,K\left(\frac{1}{2}-\xi,\,y-\eta\right)d\xi = K_F. \quad (16)$$

The function $K_F$ can be expressed in terms of elliptic integrals and has been tabulated by Weissinger.

*Evidently, once the approximation (15) has been made it is only consistent to make a corresponding approximation in (16).* This leads to the so-called $L$-method which in Weissinger's paper is based on lifting-line considerations. The result is

$$\bar{K}(y-\eta) \approx K(1,\,y-\eta) = K_L(y-\eta). \quad (17)$$

Calculations have shown[1] that results based on either $K_F$ or $K_L$ are practically indistinguishable. One advantage of $K_L$ besides its simplicity is the fact that an analogous result may be obtained without too much difficulty for wings with sweepback.[1]

With the approximation (17) the integral equation (12) may be written in the following form

$$\int_{-1}^{1}\sqrt{\frac{1+x}{1-x}}\,w\,dx$$
$$= l(y) + \frac{1}{2}\int_{-s}^{s}\frac{dl}{d\eta}\,\frac{1+\sqrt{1+(y-\eta)^2}-|y-\eta|}{y-\eta}\,d\eta. \quad (18)$$

Equation (18) shows somewhat more clearly than the corresponding result in Weissinger's paper the relation between this theory and the original lifting-line theory. It also shows, since $[1+(y-\eta)^2]^{1/2} >$

$|y - \eta|$, that the improved theory leads to more pronounced aspect ratio effects than the original theory.

We may note that with

$$l(y) = \sqrt{1 - y^2} \sum l_n y^n \tag{19}$$

the integrals on the right are expressible in terms of elliptic integrals. It is questionable, however, whether to do so will be of computational advantage.

We remark finally that here also, as in the lifting-line equations (6) and (11), it can be shown that in general the tip moments $m(\pm s)$ do not vanish.

5. *A System of Simultaneous Equations for Lift and Moment Distribution.*—We begin with the exact equation (12) and add to this the following exact equation which takes the place of the approximate equation (10)

$$\int_{-1}^{1} \sqrt{1 - x^2}\, w dx = -m(y) + \frac{1}{4} \int_{-s}^{s} \frac{dl}{d\eta} \frac{d\eta}{y - \eta} + \frac{1}{2} \int_{-s}^{s} \int_{-1}^{1} \frac{\partial p}{\partial \eta} \left\{ \frac{1}{\pi} \int_{-1}^{1} \sqrt{1 - x^2}\, K dx \right\} d\xi d\eta. \tag{20}$$

We must now take into account approximately the triple integrals both in (12) and (20). As we wish to express these triple integrals both in terms of $l$ and $m$ we can now no longer use the simple approximation (13) for $p$. Instead we set

$$\begin{aligned} p(\xi, \eta) &= [c_1(1 - \xi) + c_2(1 - \xi^2)]/\sqrt{1 - \xi^2} \\ &= -\frac{2m(\eta)}{\pi} \sqrt{\frac{1 - \xi}{1 + \xi}} + \frac{2l(\eta) + 4m(\eta)}{\pi} \sqrt{1 - \xi^2}. \end{aligned} \tag{21}$$

With (21) we obtain from (12) and (20) the following two fundamental equations

$$\int_{-1}^{1} \sqrt{\frac{1 + x}{1 - x}}\, w dx = l(y) + \int_{-s}^{s} \frac{dl}{d\eta} \left[ \frac{1}{2(y - \eta)} + K_{ll}(y - \eta) \right] d\eta + \int_{-s}^{s} \frac{dm}{d\eta} K_{lm}(y - \eta) d\eta \tag{22}$$

$$\int_{-1}^{1} \sqrt{1 - x^2}\, w dx = -m(y) + \int_{-s}^{s} \frac{dl}{d\eta} \left[ \frac{1}{4(y - \eta)} + K_{ml}(y - \eta) \right] d\eta + \int_{-s}^{s} \frac{dm}{d\eta} K_{mm}(y - \eta) d\eta. \tag{23}$$

The kernel functions $K_{ll}$, etc., are seen to be of the form

$$K_{ll} = \frac{1}{\pi^2} \int_{-1}^{1} \int_{-1}^{1} \sqrt{\frac{1 + x}{1 - x}} \sqrt{1 - \xi^2}\, K dx d\xi \tag{24}$$

$$K_{lm} = \frac{1}{\pi^2}\int_{-1}^{1}\int_{-1}^{1}\sqrt{\frac{1+x}{1-x}}\left[2\sqrt{1-\xi^2} - \sqrt{\frac{1-\xi}{1+\xi}}\right] K dx d\xi \quad (25)$$

$$K_{ml} = \frac{1}{\pi^2}\int_{-1}^{1}\int_{-1}^{1}\sqrt{1-x^2}\sqrt{1-\xi^2}\, K dx d\xi = 0 \quad (26)$$

$$K_{mm} = \frac{1}{\pi^2}\int_{-1}^{1}\int_{-1}^{1}\sqrt{1-x^2}\left[2\sqrt{1-\xi^2} - \sqrt{\frac{1-\xi}{1+\xi}}\right] K dx d\xi = -K_{ll}. \quad (27)$$

The two functions $K_{ll}$ and $K_{lm}$ may be tabulated, just as the function $K_F$ of equation (16) has been tabulated and just as the function $\tilde{K}$ in (14) may be tabulated. Once this has been done one may solve equations (22) and (23) by means of series of the form

$$l = \sum l_k \sin k\phi, \qquad m = \sum m_k \sin k\phi \quad (28)$$

where

$$\cos\phi = y/s. \quad (29)$$

Equations (22) and (23) then become systems of simultaneous equations for the coefficients $l_k$ and $m_k$. For specific applications it will be necessary to tabulate the functions

$$G_{ll}^{k}(s, \cos\phi) = \int_0^{\pi} \cos k\theta K_{ll}[s(\cos\phi - \cos\theta)]d\theta \quad (33)$$

and corresponding expressions for $K_{lm}$, for various values of $k$ and $s$, whence the usual collocation procedure can be applied; or to tabulate the corresponding quantities which occur in an application of the Multhop procedure to equations (22) and (23).

Considerable simplification is introduced into the foregoing results if again as in going from (15) to (16) and (17) use is made of the following formulae for approximate numerical integration

$$\left.\begin{array}{c}\dfrac{1}{\pi}\displaystyle\int_{-1}^{1}\sqrt{\frac{1+x}{1-x}}f(x)dx \approx f\left(\frac{1}{2}\right), \quad \dfrac{1}{\pi}\displaystyle\int_{-1}^{1}\sqrt{\frac{1-x}{1+x}}f(x)dx \approx f\left(-\frac{1}{2}\right) \\ \dfrac{1}{\pi}\displaystyle\int_{-1}^{1}\sqrt{1-x^2}f(x)dx \approx \frac{1}{2}f(0).\end{array}\right\} \quad (34)$$

Equations (24) and (25) then become

$$K_{ll} \approx \frac{1}{2}K\left(\frac{1}{2}, y-\eta\right) = \frac{\sqrt{1/4 + (y-\eta)^2} - |y-\eta|}{y-\eta} \quad (35)$$

$$K_{lm} \approx K\left(\frac{1}{2}, y - \eta\right) - K\,(1, y - \eta)$$
$$= \frac{\sqrt{1 + 4(y - \eta)^2} - \sqrt{1 + (y - \eta)^2} - |y - \eta|}{y - \eta}. \qquad (36)$$

More accurate approximations to $K_{ll}$ and $K_{lm}$ may be obtained if instead of equations (34) the following approximate numerical integration formulae are used

$$\left.\begin{aligned}
\frac{1}{\pi}\int_{-1}^{1}\sqrt{\frac{1 + x}{1 - x}}f(x)dx &\approx \frac{5 + \sqrt{5}}{10}f\left(\frac{1 + \sqrt{5}}{4}\right) + \frac{5 - \sqrt{5}}{10}f\left(\frac{1 - \sqrt{5}}{4}\right)\\
\frac{1}{\pi}\int_{-1}^{1}\sqrt{\frac{1 - x}{1 + x}}f(x)dx &\approx \frac{5 + \sqrt{5}}{10}f\left(-\frac{1 + \sqrt{5}}{4}\right)\\
&\qquad + \frac{5 - \sqrt{5}}{10}f\left(-\frac{1 - \sqrt{5}}{4}\right)\\
\frac{2}{\pi}\int_{-1}^{1}\sqrt{1 - x^2}f(x)dx &\approx \frac{1}{2}f\left(\frac{1}{2}\right) + \frac{1}{2}\int\left(-\frac{1}{2}\right).
\end{aligned}\right\} \qquad (37)$$

While equations (34) are exact when $f(x)$ is any linear function of $x$, equations (37) are exact when $f(x)$ is any third-degree polynomial in $x$.

We will omit listing the approximations to $K_{ll}$ and $K_{lm}$ which replace (35) and (36) when (37) is used.

We may note that when $m = -l/2$, which would follow from the assumption that the $(1 - x^2)^{1/2}$-term in (21) is negligible, equation (22) contains the kernel $[1/2(y - \eta)] + K_{ll} - {}^1/_2K_{lm}$. As

$$K_{ll} - \frac{1}{2}K_{lm} = \frac{1}{2\pi^2}\int_{-1}^{1}\int_{-1}^{1}\sqrt{\frac{1 + x}{1 - x}}\sqrt{\frac{1 - \xi}{1 + \xi}}\,Kdxd\xi \qquad (37)$$

equation (22) is then reduced to a combination of equations'(12) and (14) of the Weissinger procedure, which is as it should be.

It is to be expected that the solution of equations (22) and (23) is nearer to the exact solution of lifting surface theory than is the solution of the combined equations (12) and (14).

An example where the system of two simultaneous equations will give a meaningful answer whereas the one-equation procedure will not is given by a wing formed such that $w$ = const. $\partial h/\partial x = C(y)(1 - 2x)$, that is, a wing for which according to the two-dimensional theory every section has zero lift but non-zero moment. The one-equation procedure predicts that the effect of finite span does not modify the zero lift property at every spanwise station. It says nothing about the effect of finite span on the moment distribution. The present two-equation procedure permits determination of the spanwise variation of moment distribution and indicates that the effect of finite span modifies the zero lift properties which the two-dimensional theory predicts.

It is thought that the idea underlying the proposed two-equation procedure, which may be extended to an $n$-equation procedure, will also be useful in other problems of lifting-surface theory.

*Summary.*—The steady motion of a rectangular lifting surface in incompressible flow is considered, with the following two objectives. (1) A modified derivation is given of Weissinger's improvement of Prandtl's lifting-line theory. (2) The integral equation of lifting-line theory (and Weissinger's improvement) for the spanwise variation of lift intensity is generalized to two simultaneous integral equations for lift intensity and moment intensity. In so doing certain difficulties of lifting-line theory with reference to behavior at and near tip sections are substantially ameliorated. The two-equation procedure of this paper is shown to be applicable to a problem which is outside the scope of the lifting-line equation.

---

[*] This paper is based on an investigation which was carried out for Chance Vought Aircraft, Division of United Aircraft Corporation, under sponsorship of the Bureau of Aeronautics, U.S. Navy.

[1] Weissinger, J., "The Lift Distribution of Swept Back Wings," NACA T.M. No. 1120, March, 1947 (Translation of German ZWB F. B. 1553 (1942)).

[2] See for instance reference 1.

---

# RESPONSES OF THE PROXIMAL RETINAL PIGMENT OF THE ISOLATED CRUSTACEAN EYESTALK TO LIGHT AND TO DARKNESS

By L. H. Kleinholz

Department of Biology, Reed College, Portland, Oregon

Communicated by G. H. Parker, February 15, 1949

The possibility of an endocrine factor in regulating the photomechanical changes of the pigment cells of the crustacean retina was demonstrated by Kleinholz (1934, 1936) who showed that extracts of the eyestalks of *Palaemonetes* when injected into dark-adapted individuals brought about a movement of the retinal pigments to the position characteristic of the light-adapted retina. These hormonal effects have since been confirmed by Welsh (1939) for *Cambarus* and by Nagano (1947) for *Paratya.* Smith (1948) has recently reported similar effects on the retina of brachyuran crustaceans but has not detailed the types of pigment cells concerned since his observations were made indirectly on the basis of disappearance of "glow" after such injections. It was further shown by Welsh (1941) and by Smith (1948) that the sinus gland within the eyestalk is the major source of the activity affecting these retinal pigments.

Some differences in the response of the retinal pigments to injected eyestalk extract have been observed in the different species of crustaceans investigated. The prawns studied by Kleinholz and by Nagano showed movement of the distal and of the reflecting pigments into the light-adapted position after injection of eye-stalk extract into dark-adapted individuals; the proximal pigment, however, remained dark adapted. In *Cambarus* Welsh (1939) found that the proximal pigment had a higher threshold than the distal pigment for the injected eye-stalk extract, because when more concentrated extracts were used, the proximal pigment also moved into the light-adapted position.

As has already been pointed out (Kleinholz, 1942, 1947) conclusive proof of endocrine regulation of the photomechanical changes of the retinal pigment cells would be demonstrated, according to conventional endocrine procedure, by experiments complementing those described above. That is, if injection of sinus gland extract causes light adaptation of a retinal pigment, then surgical removal of the sinus gland should result in the dark-adapted phase of that pigment, or should result in that pigment's being unable to move into the light-adapted position under conditions which ordinarily bring about light adaptation.

In attempting to resolve this point regarding the proximal pigment of the crayfish, experimental results were obtained that indicated the activity of this retinal pigment was more complicated than was originally suspected, and that possibly more than a single mechanism of control may be concerned in regulating its activity. The effect of surgical removal of the sinus glands from the eyestalks of the crayfishes, *Cambarus* and *Astacus*, on the ability of the proximal pigment to respond to light and to darkness has been described in preliminary reports (Kleinholz, 1947, 1948). In these animals it was demonstrated that apparently normal light and dark adaptation of the proximal retinal pigment occurred after sinus gland removal. The conclusion reached in the 1948 account was that, in addition to regulation of the proximal pigment by the sinus gland, other regulatory sources of endocrine secretion outside of the sinus gland might be present, as well as the possibility that the proximal pigment cells of the crayfish retina could respond directly to light and to darkness.

The experiments reported here were performed with four groups of the crayfish, *A. trowbridgei*. Two of these groups were initially dark adapted in a photographic dark room, while the remaining two groups were initially light adapted by keeping the animals in a brightly illuminated environment. In the first of these light-adapted groups, one eyestalk was carefully ablated from each animal in the group, placed in a moist chamber and the isolated stalks placed in darkness; the second eyestalk of each animal in the group was then excised and fixed, to verify the degree of light adaptation of the retinal pigments at the time of ablation of the experimental eyestalk. After the isolated eyestalks had been in darkness for 3 to 5 hours they were fixed. Both the experimentally isolated eyestalks and the control eyestalks were subsequently sectioned and the position of the pigment cells examined histologically. The proximal pigment of the isolated retinas was in the position typical for dark adaptation, while that of the control eyestalks was typically light adapted.

The second group of light-adapted crayfish was used as a control for the procedures employed with the first group. In this group the isolated eyestalks were placed in moist chambers and kept on an illuminated white background; the second eyestalk of each animal was fixed as before. The ablated eyestalks were fixed after illumination for 3 to 5 hours. Examination of the sectioned retinas showed that the proximal pigments of both sets were completely light adapted.

Similar experiments were performed with the two groups of dark-adapted *Astacus*. In the first group one eyestalk from each animal was removed (in about 30 seconds) by red light from a photographic safelight, and, after being put in the moist chamber, was placed on an illuminated white background. The second eyestalk of each animal was immediately fixed as a control while the isolated stalks were fixed after 5 hours of illumination. Histological examination revealed the proximal pigment of the isolated eyestalks to be light adapted, while the control retinas were dark adapted. The second group of dark-adapted animals served as control for the methods used. In this set, both the isolated eyestalks (which had been put in moist chambers and maintained in darkness for about 5 hours) and the second eyestalk of each animal showed the proximal pigment in the position characteristic for the dark-adapted retina.

The evidence of the above experiments indicates that proximal retinal pigment cells can respond directly to light and to darkness while isolated from the central nervous system (except for the optic ganglia) and from the circulatory system. A more detailed account of these and of other experiments indicating the rôle of the nervous system in retinal pigment migration will be reported later.

Kleinholz, L. H., "Eye-Stalk Hormone and the Movement of the Distal Retinal Pigment in *Palaemonetes*," PROC. NAT. ACAD. SCI., **20**, 659–661 (1934).

Kleinholz, L. H., "Crustacean Eye-Stalk Hormone and Retinal Pigment Migration," *Biol. Bull.*, **70**, 159–184 (1936).

Kleinholz, L. H., "Hormones in Crustacea," *Biol. Rev.*, **17**, 91–119 (1942).

Kleinholz, L. H., "Migrations of the Retinal Pigments and Their Regulation by the Sinus Gland," Confer. scientifique internat. sur l'endocrinologie des arthropodes. Centre National de la Recherche Scientifique, Paris (1947).

Kleinholz, L. H., "Factors Controlling the Migration of the Proximal Pigment of the Crustacean Retina," *Anat. Record*, **101** (4), 15 (1948).

Nagano, T., "Physiological Studies on the Pigmentary System of Crustacea. II. The Pigment Migration in the Eyes of the Shrimps," *Science Repts. Tôhoku Univ., Fourth Ser. (Biol.)*, **18**, 1–16 (1947).

Smith, R. I., "The Role of the Sinus Glands in Retinal Pigment Migration in Grapsoid Crabs," *Biol. Bull.*, **95**, 169–185 (1948).

Welsh, J. H., "The Action of Eye-Stalk Extracts on Retinal Pigment Migration in the Crayfish, *Cambarus bartoni*," *Ibid.*, **77**, 119–125 (1939).

Welsh, J. H., "The Sinus Glands and 24-Hour Cycles of Retinal Pigment Migration in the Crayfish," *J. Exp. Zoöl.*, **86**, 35–49 (1941).

---

# *THE SURFACE GEL LAYER OF FUNDULUS EGGS IN RELATION TO EPIBOLY*

BY J. P. TRINKAUS

OSBORN ZOOLOGICAL LABORATORY, YALE UNIVERSITY, AND MARINE BIOLOGICAL LABORATORY

Communicated by R. G. Harrison, February 3, 1949

There have been several studies relating the properties of the surface gel layers of Amphibian[1] and Teleost[2] eggs to the movements of gastrulation. Teleost eggs, with their remarkable epiboly, offer excellent material for a further analysis of this question. The present report outlines certain preliminary studies on the properties of the surface gel layer of the eggs of *Fundulus heteroclitus*, with special reference to its possible rôle in epiboly.

All studies were made on eggs from which the chorion (or "shell") had been removed. Unless otherwise stated, unbuffered double-strength Holtfreter's solution was the medium employed. It was possible to demonstrate at the outset that Fundulus eggs possess a surface gel layer, by microdissection with glass needles. This gel layer constitutes the outer coat in oviducal eggs, unfertilized eggs and in eggs activated by sea water, as well as in developing eggs at all stages as far as closure of the blastopore. It is continuous, non-pigmented and constitutes the outer layer of both the unsegmented yolk sphere and the blastodisc. The outer surface of this gel layer is non-adhesive. If a blunt instrument is pressed against the egg, the gel layer shows resistance and is thrown into radiating folds. Microdissection of all stages through blastopore closure (Oppenheimer, stage-15)[3] indicates that the surface gel layer is the outer layer of the epiblast to which all epiblast cells are attached. Thus the epiblast tends to come off in sheets when removed. Furthermore, the gel layer of the epiblast is continuous with that of the yolk.

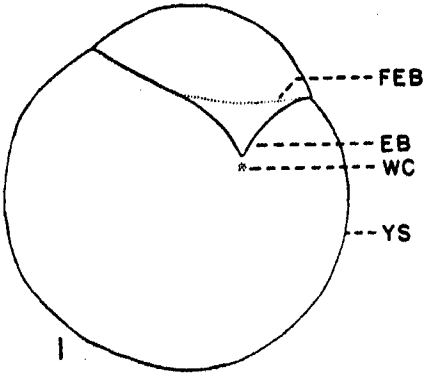
*Properties of the Surface Gel Layer.*—When a wound is made in the gel layer of the yolk sphere, the aperture first becomes round and expands in size as the gel layer retracts, presumably because of the tension resulting from the turgor of the yolk. This rounding up and expansion of the wound also indicates that the gel layer possesses elastic properties. After this initial increase in size the aperture is closed rapidly, while the relatively fluid yolk continues to ooze during all but perhaps the final stages of the process. This whole process takes from 2–5 minutes depending upon the size of the wound. As the wound closes, folds or stress lines appear in the gel layer, radiating from the wound. The gel layers of oviducal eggs and unfertilized eggs activated by sea water have the same ability to close wounds rapidly. This process takes place in typical fashion when the eggs are in single, double or quadruple-strength Holtfreter's solution, in sea water and in distilled water.

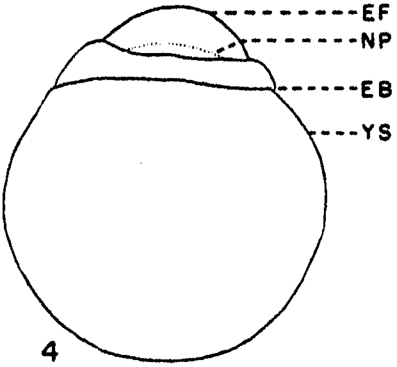
In order to ascertain whether wound closure is actually due to contraction of the surface gel layer, carbon marks[4] were placed on opposite edges of an open wound (prior to the onset of wound closure). As the wound closed, the marks were drawn toward each other until they almost met at the point of wound closure. Such an experiment indicates that wound closure is chiefly the result of the contraction of the edges of an already existing surface layer, rather than the formation of a new membrane from the yolk cytoplasm. The very last stage appears to be a process of protoplasmic "clotting," wherein a small amount of yolk cytoplasm coagulates and forms a plug.

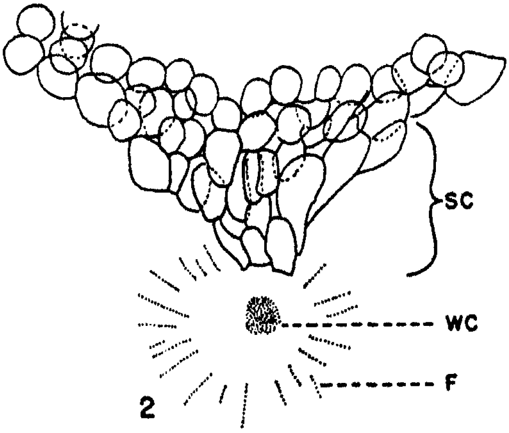
If carbon marks are made at various points on the surface of one hemisphere of the egg and a wound then made in the midst of these marks those marks nearest the closing wound move toward each other but the distance between these marks and marks further away increases. Thus the surface layer stretches or expands as the process of wound closure proceeds. It seems also from this experiment that a large part, if not the total, surface layer of the egg is involved. The integrity of the entire surface layer is not necessary for wound closure, however, inasmuch as two or more wounds in the surface layer will close at the same time, in typical fashion, even though one of the wounds is quite large, e.g., 1 mm. in diameter, which is one-half the diameter of the entire egg.

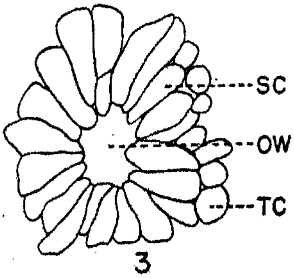
Wound closure in the surface gel layer of the yolk sphere near the edge of the blastoderm, at all stages from the one-celled blastodisc to the closure of the blastopore, results in a stretching of the blastoderm in the direction of the closing wound, and also in a stretching of individual adjacent cells and periblast nuclei as well. This phenomenon (see Figs. 1 and 2), takes place in a few minutes' time. It demonstrates that the cells at the edge of the blastoderm (and nuclei of the marginal periblast) are attached to the surface gel layer of the yolk sphere in such a manner that tension in this layer exerts a corresponding tension on them, resulting in an appropriate distortion of cell shape.

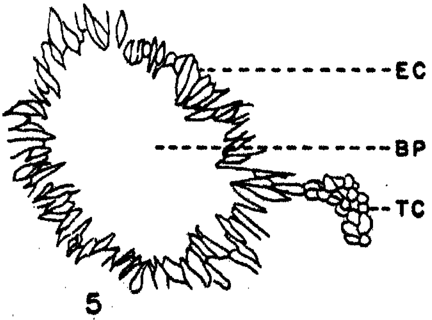
Experimental demonstration that the surface gel layer is continuous


1


4


2


3


5

over the entire surface of the blastoderm is provided by further wound healing experiments. If a group of cells is removed from the epiblast layer in a blastula or gastrula stage, a process of wound closure is initiated, which appears similar to that in the yolk surface layer, although it is slower, taking 7 to 12 minutes or longer. It results in a radial stretching of the cells adjacent to the wound. This stretching becomes apparent when the wound is approximately half closed, and is continually augmented until the completion of the process (see Fig. 3). Such a radial stretching effect is clear evidence that a supra-cellular force, presumably the surface gel layer, is concerned in wound closure in the epiblast cell layer.

A wound in the surface of one of the blastomeres in an early cleavage stage is rapidly closed in a similar manner, with resulting radial folds or lines of stress in the blastomere surface. Apropos of wound healing in early blastomeres, it might be mentioned that it is relatively difficult to wound the surface of a blastomere of the 1- or 2-cell stage. The surface seems tougher than that of the yolk. Furthermore, since little or no cytoplasm pours out of the wound, the cytoplasm of these cells must have a higher viscosity than the yolk.

As a result of these various wound healing experiments it is evident that the outer surface layer of Fundulus eggs is a gel layer possessed of marked elastic properties. In this respect the surface gel layer of Fundulus eggs is like that of Amphibian and trout eggs (cf. Holtfreter[1] and Devillers[2]). However, if rapidity of wound closure be an adequate criterion,

←

EXPLANATION OF PLATE

(Drawings based on camera lucida tracings)

FIGURE 1

Stretched edge of blastoderm resulting from wound closure in surface gel layer of yolk sphere in Oppenheimer stage 10. × 20.

FIGURE 2

Stretched cells at the edge of the blastoderm resulting from wound closure in the surface gel layer of the yolk sphere. Stage 8. × 100.

FIGURE 3

Stretched cells of the epiblast after partial wound closure. Stage 11. × 220.

FIGURE 4

Protrusion of exposed floor of blastocoel, after removal of roof of blastocoel. Stage 10+. × 20.

FIGURE 5

Elongate cells at lips of blastopore in normal embryo. Stage 15. × 100.

*BP*, blastopore; *EB*, edge of blastoderm; *EC*, elongate cells at lips of blastopore; *EF*, exposed floor of blastocoel; *F*, radiating folds in surface gel layer; *FEB*, former edge of blastoderm; *NP*, normal position of floor of blastocoel; *OW*, open wound; *SC*, stretched cell; *TC*, typical epiblast cell; *WC*, point of wound closure; *YS*, yolk sphere.

the elasticity of the Fundulus egg gel layer is considerably greater than that of Amphibian and trout eggs.

The floor of the segmentation cavity or blastocoel was studied in late cleavage and blastula stages to see if it likewise possesses such an elastic surface layer. When a large section of the roof of the blastocoel is removed, the floor of the blastocoel is exposed, consisting at this stage solely of the surface of the yolk sphere. It is not yet covered by periblast. Within a few minutes after such an operation, the exposed surface bulges out as shown by figure 4. This demonstrates both the turgor of the yolk, and the lack of elastic strength in the surface of the yolk at the floor of the segmentation cavity. When a wound is made in the yolk surface at the floor of the segmentation cavity it gapes and remains open. The relative ease with which such a wound is made itself indicates the weakness of this layer. It is thus apparent that the yolk surface at the floor of the segmentation cavity in early stages lacks a surface layer with the same elastic, contractile properties as are possessed by the peripheral surface of the egg. Similar wounding experiments on the floor of the blastocoel in late blastulae and all stages of gastrulae, when the floor is covered with periblast, suggests that the surface at these stages possesses an elastic gel layer, inasmuch as wound closure seems more typical, although often slower (ca. 10 min.).

If an early blastula with an open wound in the floor of the blastocoel remains in the bottom of a dish for an hour or so, all of the yolk oozes out, and all that remains is the surface gel layer lying in the bottom of the dish with marginal blastoderm cells clinging to the edges of the aperture. It has the appearance of an empty cellophane bag. This forms superb material for physicochemical studies. A few preliminary experiments have been done along these lines.

*Mechanical and Chemical Studies.*—The gel layer, in this isolated condition, will remain intact and apparently in a normal state for at least 12 hours in double-strength Holtfreter's solution. It is non-adhesive on its outer surface (as in the intact egg) and sticky and viscous on its inner surface. If the inner surfaces of 2 pieces are pressed together they will adhere quite firmly. The elastic properties of the material can be demonstrated by stretching pieces between two glass needles. In this manner a piece may be stretched to half again its length, returning to its original length upon release. Small pieces of the gel layer, with cells clinging to them, tend to round up into spherical globules. When the gel layer is emptied of yolk and, therefore, in a slack condition, a wound in it no longer expands into a round opening in the usual manner. Apparently, the normal process of wound closure in the surface gel layer is dependent upon a state of tension in the gel layer, resulting from the normal turgidity of the egg. After a certain small amount of mechanical manipulation, the isolated gel layer completely solates. This is a property of thixotropic gels. It does not return again to the gel state, however, upon standing, as is typical of these gels.

In $Ca^{++}$-free sea water the surface gel layer of the intact egg dissolves, or better—solates, after the egg has been in the medium for 1–2 hours. In some cases the egg first flattens out on the bottom of the dish like a thick pancake, suggesting a general thinning and weakening of the gel layer, prior to solation. In other cases, the egg retains its spherical shape and its normal wound healing ability for approximately $1^1/_2$ hours. At about this time, however, a wound forms readily without extreme intervention and remains open. All of the surface gel layer then solates in the course of a few minutes. At no time in any of these cases was the outer surface observed to become sticky, as is normal for the inner surface. Both development and wound closure of Fundulus eggs take place normally in distilled water. The effect of $Ca^{++}$-free sea water is, therefore, not due simply to a lack of $Ca^{++}$-ions. The positive action of the monovalent ions in the ionically unbalanced solution must also be considered.

The behavior of the surface gel layer of the Fundulus egg in $Ca^{++}$-free sea water is characteristic of the behavior of the surface gel layers of other eggs in $Ca^{++}$-free media, e.g., Amphibian. The hyaline plasma layer of Echinoderm eggs also solates in $Ca^{++}$-free sea water. This latter fact suggests that the similarities of the surface gel layer and the hyaline plasma layer may be more than superficial, a subject which invites further investigation.

*Possible Role of Surface Gel Layer in Epiboly.*—The properties of the surface gel layer, as outlined in the previous pages, suggest that it may be of some importance in relation to the movements of gastrulation. It certainly seems probable that the surface gel layer, being the outer layer of the epiblast to which all epiblast cells are attached, acts as a coördinating and unifying force in the highly integrated cell movements characteristic of the epiblast in Teleost gastrulation. Furthermore, the stretching effect of the gel layer upon the edge of the blastoderm in wound healing is especially suggestive of a possible causal rôle of the gel layer in epiboly. To test this idea, detailed observations were made of cell shape at the edge of the blastoderm in all stages of epiboly (in *normal* eggs). At the stage when epiboly *seems* to proceed at the greatest rate, namely the final closure of the blastopore (stage-14+ to stage-15), there is a radial elongation of the cells at the lips of the blastopore (see Fig. 5). These elongate cells appear to be stretched toward the point of blastopore closure. This suggests that the blastoderm is being pulled over the surface of the yolk by the contractile tension of the surface gel layer of the yolk sphere. If this be the case, one should be able to cause a uniform stretching of cells at the lips of the blastopore in an earlier stage, e.g., stage-13+ or stage-14, by making a wound in the surface gel layer at the point of future blastopore closure. When this was done, cells at the edges of the blastoderm became stretched in the direction of the point of closure of the blastopore. Furthermore, it was possible by such an experiment to cause precocious closure of the blastopore with the embryo only in an embryonic shield stage (stage-13+).

One further case supports this possible rôle of the surface gel layer in epiboly. In the series of experiments in which the roof of the blastocoel was removed during a late blastula stage (stage-11) one egg remained intact over an 18-hour period. In spite of the fact that the large wound in the roof of the blastocoel remained open, epiboly took place completely with resulting closure of the blastopore. Such a result constitutes supporting evidence for the thesis that epiboly is not an overgrowth, as has often been thought, but is rather the result of the blastoderm being pulled down over the yolk.

If the hypothesis is true that epiboly is due to the contractile tension of the surface gel layer of the yolk sphere on the cells of the blastoderm, it is likely that in the course of epiboly there would be a greater stretching of the blastoderm nearest its edge, with gradually decreasing amount of stretching as one proceeds toward the animal pole, assuming that the surface gel layer of the blastoderm has uniform elasticity. It is certainly true, for example, that, after the closing of a wound in the surface gel layer of the yolk sphere, only the marginal blastoderm cells, nearest the closing wound, show marked stretching.

To check this point, the relative degree of stretching of different regions of the blastoderm was studied by the method of carbon marking. Marks were placed at various points on the surface of the epiblast in an early gastrula stage. Camera lucida records of the positions of these marks relative to one another were made every few hours during the course of epiboly. The distance between carbon marks over the entire blastoderm increased uniformly, as epiboly proceeded. This sort of result suggests that there is a uniform expansion, or spreading, of the epiblast, and, therefore, also of the surface gel layer of the blastoderm, in epiboly. Such a phenomenon indicates that epiboly may not be simply the result of tensions exerted on the edge of the blastoderm by the surface gel layer of the yolk sphere. The situation is more complex, it would seem, than some previous workers have been willing to admit (e.g., Lewis, who, however, quotes no experimental evidence). It is, of course, possible that any tension exerted on the edge of the blastoderm during epiboly is too small to be detected by the carbon marking technique. Nevertheless, these data indicate that the expansion of the surface gel layer of the epiblast must also be considered as a possible causal factor in epiboly.

In spite of these suggestive observations and experimental results, it is still an open question as to whether the surface gel layer actually plays a determining rôle in epiboly in the Fundulus egg. If it is causally important in epiboly, the precise rôle it plays has not yet been determined. Further investigation of the relation of the surface gel layer to the cell movements of early morphogenesis is necessary, for it may possibly shed some light on one of the more obscure aspects of causal embryology, the mechanics of gastrulation.

It should be emphasized that this report has been concerned exclusively with studies on eggs of *F. heteroclitus*. The results are not necessarily representative of Teleost eggs in general. In fact, the surface layers of the eggs of some forms, e.g., *Platypoecilus maculatus* and *Lebistes reticulatus*, seem to have a much lower degree of elastic strength than that of Fundulus eggs. The surface layer of Toadfish eggs (*Opsanus tau*) and Medaka eggs (*Oryzias latipes*), however, appear to be more like that of the Fundulus egg. The properties of the surface gel layer suggest that one of its functions may be to protect the egg from the effects of osmotic changes (as pointed out by Devillers) and mechanical buffeting. The weakness of the gel layers of Lebistes and Platypoecilus eggs thus may be related to the viviparity of these forms, and, conversely, the strength of the gel layers of Fundulus, Opsanus and Oryzias eggs to oviparity.

*Summary*.—Fundulus eggs possess a surface gel layer, which constitutes the continuous, outer layer of both the unsegmented yolk sphere and the blastoderm. It has striking elastic properties and plays a decisive rôle in wound healing. The contraction of the surface gel layer in wound closure stretches adjacent cells and pulls a large sector of the blastoderm toward the wounded area. At closure of the blastopore in normal embryos, the cells at the blastopore lips elongate. At an earlier stage, this phenomenon is simulated and blastopore closure hastened if the surface gel layer is wounded at the point of future closure of the blastopore, suggesting that epiboly may be the result of contractile tension exerted by the surface gel layer of the yolk sphere. The behavior of the blastoderm, as observed in marking experiments, indicates that the spreading of the surface gel layer of the blastoderm must also be considered as a possible factor in causing epiboly.

[1] Holtfreter, J., *J. Exp. Zool.*, **93**, 251–323 (1943); *Ibid.*, **94**, 261–318 (1943); **95**, 171–212 (1944); *Ann. N. Y. Acad. Sci.*, XLIX, 709–760 (1948).

[2] Lewis, W. H., *Anat. Rec.*, **85**, 326 (1943); *Biol. Bull.*, **87**, 154 (1944). Devillers, Ch., *C. R. Acad. Sci.*, **226**, 1310–1312 (1948); *Ann. Station Centrale Hydrobiologie applique*, in press (1948). Trinkaus, J. P., *Biol. Bull.*, **95**, 271 (1948).

[3] Oppenheimer, Jane, *Anat. Rec.*, **68**, 1–15 (1937).

[4] Spratt, N. T., Jr., *J. Exp. Zool.*, **103**, 259–304 (1946).

PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES

Volume 35 May 15, 1949 Number 5

# *THE PREPARATION OF SPECIFIC ADSORBENTS*

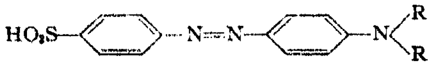
BY FRANK H. DICKEY*

DEPARTMENT OF CHEMISTRY, CALIFORNIA INSTITUTE OF TECHNOLOGY†

Communicated by L. Pauling, March 21, 1949

A method has been developed for preparing adsorbents having specific affinities for predetermined substances. This method consists in forming the structure of the adsorbent in the presence of the particular compound for which it is desired to prepare a specific adsorbent. The presumable explanation of the formation of a specifically attracting structure under these conditions is that the adsorbent in the process of formation has accessible to it a very great number of structures which differ only slightly in stability, and that in the presence of a foreign molecule those structures that are stabilized through attraction for the foreign molecule are preferentially assumed. The adsorbent is thus pictured as automatically forming pockets that fit closely enough to the foreign molecule to hold it by van der Waals' forces, hydrogen bonds, interionic attractions, and other types of intermolecular interaction. This mechanism is the same as that proposed by Pauling[1] for the formation of antibodies with use of antigen molecules as a template, which formed the basis for the manufacture of artificial antibodies reported by Pauling and Campbell.[2]

An experimental test of the idea has been made by the formation of silica gel in the presence of methyl orange and some of its homologs—ethyl, *n*-propyl and *n*-butyl orange. These compounds have the following structure:



where *R* represents the radicals named. The adsorbents were prepared by mixing 0.5 g. of finely divided dye with 30 ml. of aqueous sodium silicate ($d^{20}$ 1.401) 275 ml. of water and 30 ml. of glacial acetic acid, drying (at room temperature), grinding and sieving the resultant gels, and subjecting the fraction between 48 and 200 mesh to methanol extraction, which removed most of the dye. Control gels, without dye, were prepared in the same way.

The gels were tested as adsorbents by bringing 1-g. samples into equilibrium with 10-g. portions of standard solutions of the dyes in 5% acetic acid. The initial dye concentrations were $1.5 \times 10^{-5}$ formal. The final concentrations in the supernatant solutions were measured with a spectrophotometer and the quantity adsorbed was calculated by difference. In order to obtain accurate estimates of the dye concentrations from the spectrophotometer readings it was necessary to correct these readings for the cloudiness which appeared in the solutions after contact with the silica gel samples. Such corrections were calculated on an empirical basis from readings taken in another portion of the spectrum where the dye absorptions are negligible. The ratio of the concentration in the gel to that in the supernatant solution was then calculated as a measure of the "adsorption power" of the gel for the particular dye. The following results were obtained.

TABLE 1

| | ADSORPTION POWER FOR: | | | |
|---|---|---|---|---|
| | METHYL ORANGE | ETHYL ORANGE | PROPYL ORANGE | BUTYL ORANGE |
| Control gel | 84 | 80 | 240 | 320 |
| Specific adsorbent | 300 | 740 | 5000 | 5000 |

In order to show that this process has effected a specific rather than a general enhancement of the adsorption properties of the gels, adsorption powers were determined with each of the gels for each of the dyes. The ratio of these adsorption powers to the adsorption powers shown for the particular dyes by the control gels were then calculated. These "relative adsorption powers" are tabulated below.

TABLE 2

| | RELATIVE ADSORPTION POWER FOR: | | | |
|---|---|---|---|---|
| | METHYL ORANGE | ETHYL ORANGE | PROPYL ORANGE | BUTYL ORANGE |
| Gel prepared with: | | | | |
| Methyl orange | 3.5 | 1.6 | 1.1 | 1.1 |
| Ethyl orange | 2.5 | 9 | 2.1 | 2.2 |
| Propyl orange | 2.3 | 5 | 20 | 6 |
| Butyl orange | 1.5 | 2.8 | 5 | 15 |

It is difficult to reproduce exactly the adsorption properties of silica gel samples and these data are to be regarded primarily as a qualitative demonstration of the preparation of specific adsorbents. However, the essential features of the above table, maximum adsorption power for the dye used in the preparation of the adsorbent, and sharply decreasing adsorption powers for the dyes of decreasing similarity of molecular structure, have been observed in several independent experiments. Gels prepared with this series of compounds are not better adsorbents than the control gels for a structurally unrelated dye, fluorescein. The results obtained with this series of compounds are especially striking because the dyes differ only in the nature of the alkyl groups. Still more pronounced effects would be expected in studies of compounds having greater structural differences.

Many obvious applications and extensions of these ideas are now being investigated and detailed results will be published later. Of particular interest are an investigation of the possibility that optical isomers can be separated with such adsorbents and an attempt to demonstrate "specific catalysis," analogous to the action of enzymes, by specific adsorbents for the reactants or products of various reactions.

---

* E. I. du Pont Predoctoral Fellow in Chemistry.

† Contribution No. 1282 from the Gates and Crellin Laboratories of Chemistry, California Institute of Technology, Pasadena, California.

[1] Pauling, L., *J. Am. Chem. Soc.*, **62**, 2643–2657 (1940).

[2] Pauling, L., and Campbell, D. H., *J. Exptl. Med.*, **76**, 211–220 (1942).

---

# *THE VALENCE-STATE ENERGY OF THE BIVALENT OXYGEN ATOM*

BY LINUS PAULING

CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA*

Communicated March 21, 1949

A significant simplification in the discussion of the structure of molecules can be achieved by referring the energies of molecules not to the energies of their constituent atoms in any one of the stationary spectroscopic states of the isolated atoms, but instead to a hypothetical valence state of each atom.[1] The valence state of an atom is defined as that state in which it has the same electronic structure as it has in the molecule. Thus in the valence state the spins of the electrons that are used for forming bonds in a molecule are considered to resonate between positive and negative values, corresponding to the resonance of each bonding electron with another electron of opposite spin in the atom with which the bond is formed. For the bivalent oxygen atom, with six $L$ electrons, two orbitals, one mainly $2s$ in character and one $2p$, are occupied by unshared pairs, and the other two orbitals, mainly $2p$ but with a small amount of $s$ hybridization, are occupied by one bonding electron apiece, the spins of which are parallel 50 per cent of the time and antiparallel 50 per cent of the time. In the following paragraphs values of the energy of this bivalent state of the oxygen atom relative to the normal state of the isolated atom, $^3P$, are derived by four different methods. It is found that the four values agree reasonably well with one another, and their average value, 0.74 v.e., is indicated to be reliable to about 0.05 v.e.

According to the Slater theory, the energy values of the three Russell-Saunders states based upon the normal configuration of the oxygen atom are $F^0 - \frac{5}{25} F^2$, $F^0 + \frac{1}{25} F^2$, and $F^0 + \frac{10}{25} F^2$, for $^3P$, $^1D$, and $^1S$, respectively. The spectrum of oxygen shows that the $^1D$ state lies 9239 cm.$^{-1}$ above $^3P$ and $^1S$ 9739 cm.$^{-1}$ above $^1D$. These differences in energy are not in the ratio 2:3 predicted by simple spectroscopic theory. For the valence state of the atom, neglecting hybridization, the energy is predicted to be equal to the Coulomb interaction of two $p$ electrons occupying different orbitals directed along two mutually orthogonal axes, $F^0 - \frac{2}{25} F^2$, and one-half of the resonance integral between these two orbitals, that is, $-\frac{3}{50} F^2$, a total of $F^0 - \frac{7}{50} F^2$. $F^2$ is indicated to have the value 6.16 v.e. by the $^1S - {}^1D$ separation, and 8.21 v.e. by the $^1D - {}^3P$ separation. If the $^3P$ state is considered to be depressed in energy through resonance with a $^3P$ state based on a higher configuration, then the energy of the valence state relative to $^3P$ is predicted to be 0.86 v.e., which is the energy of $^1D$ corrected by the quantity $-\frac{9}{50} F^2$ with $F^2$ given by the $^1S - {}^1D$ separation. Similarly if $^1S$ is considered to be depressed by resonance, the energy of the valence state is predicted to be 0.49 v.e. In the absence of knowledge about the nature of the cause of the deviation of the intervals from the theoretical ratio the best that we can do is to take the average, 0.68 v.e., as the spectroscopic value of the energy of the valence state.[2] The true value would be expected to be slightly higher because of hybridization of the bond orbitals; $s$-$p$ hybridization in the alkali-metal diatomic molecules gives rise to a valence-state energy of 0.1 to 0.2 v.e. for the alkali atoms,[3] and similar effects of hybridization may be expected for other atoms.

A second, independent value of this energy quantity can be obtained by use of the postulate that the Morse function represents correctly the bond energy of two atoms in their valence states, and that the energy of a molecule obtained by extrapolation from the low-lying vibrational energy levels corresponds to the valence states of the atoms in the molecule. Linear extrapolation of the vibrational energy levels of the normal state of the oxygen molecule, $O_2X^3\Sigma_g^-$, with use of the values 1580 cm.$^{-1}$ for $\omega_e$ and 11.35 cm.$^{-1}$ for $\omega_e x_e$ (as given by the slope of the best straight line for $\Delta G(v + {}^1/_2)$ to $v = 17$), leads to 6.69 v.e. for the energy of dissociation to two oxygen atoms in the valence state. The dissociation energy of the normal oxygen molecule into oxygen atoms in the normal state is 5.084 v.e., and in consequence the energy of the valence state of the oxygen atom is indicated to be 0.80 v.e., one-half the difference of these two quantities.

A similar calculation can be made for the OH molecule. This molecule is to be considered as containing a bivalent oxygen atom with one valence free. The extrapolated energy of dissociation into a hydrogen atom and an oxygen atom in its valence state is 5.03 v.e., corresponding to the observed values $\omega_e = 3735$ cm.$^{-1}$ and $\omega_e x_e = 82.6$ cm.$^{-1}$. The dissociation energy into a hydrogen atom and a normal oxygen atom is 4.34 v.e., as given by the work of Dwyer and Oldenberg.[4] The difference between these values, 0.69 v.e., is the energy of the valence state of oxygen in this molecule.

It is very interesting to note that the difference between the energy of removal of the first hydrogen atom from a water molecule, involving breaking one oxygen-hydrogen bond, and the energy of removal of the second hydrogen atom vanishes when the valence state of oxygen is used as the point of reference. The energy of removal of the first hydrogen atom is found by combining thermochemical data with Dwyer and Oldenberg's value for the dissociation energy of OH to be 5.12 v.e., whereas that for the second hydrogen atom is 4.34, relative to the normal state of the oxygen atom. If the valence state of oxygen is used as a point of reference, the second of these values is to be increased by the energy of the valence state, which has been shown above to be approximately 0.7 v.e., and it is thus brought into approximate equality with the first bond energy. The equality of these two energy quantities to within 2 per cent is a striking illustration of the significance of the chemist's picture of the molecule as involving chemical bonds between pairs of atoms.

If the assumption is made that in this molecule the energy of breaking the first bond is exactly equal to the energy of breaking the second bond, when the valence state of oxygen is taken as the point of reference, an independent value can be derived for the energy of the valence state of oxygen. This value, 0.78 v.e., the difference of the two dissociation energies, agrees reasonably well with the others.

The agreement between the values of the valence-state energy of the oxygen atom in three different molecules, $O_2$, OH, and $H_2O$, suggests that the same value will hold for other molecules also, and this suggestion is supported by the agreement with the range of values given by the simple theory of atomic spectroscopic states. The average of the four independent values of the energy of the bivalent state of the oxygen atom relative to the normal spectroscopic state, 0.74 v.e. or 17.1 kcal./mole, is presumably reliable to about 0.05 v.e.

---

* Contribution No. 1283 from the Gates and Crellin Laboratories of Chemistry.

[1] Pauling, L., *Sommerfeld Festschrift, Z. Naturforsch.*, **3a**, 438–447 (1948).

[2] In the discussion of the valence state of oxygen in reference 1 the energy of the valence state was given incorrectly as $F^0 - \frac{11}{100} F^2$.

[3] Pauling, L., *Proc. Roy. Soc., London*, **A196**, 343–362 (1949).

[4] Dwyer, R. J., and Oldenberg, O., *J. Chem. Phys.*, **12**, 351–361 (1944).

---

# *ACQUISITION OF THE J SUBSTANCE BY THE BOVINE ERYTHROCYTE**

BY CLYDE STORMONT[†]

DEPARTMENT OF GENETICS, UNIVERSITY OF WISCONSIN, MADISON, WISCONSIN

Communicated by R. A. Brink, April 2, 1949

The antigenic pattern of the bovine erythrocyte is genetically conditioned by numerous genes on at least nine of the thirty pairs of chromosomes.[1, 2] The combinations of these genes permit well over a million blood types. Any two animals taken at random, therefore, will almost invariably possess distinct blood types. A study of bovine twins, however, disclosed a much higher proportion of identical blood types than is expected in view of the relatively low frequency of monozygotous twinning in cattle.[3] Owen pointed out that this high proportion of identical blood types could be accounted for as a consequence of the known frequent union of chorionic blood vessels of twin fetuses, as demonstrated by Lillie.[4] Apparently, embryonal blood cells are interchanged by those twins having a common circulatory system and these primordial cells settle in the hematopoietic tissues where they serve as a source of erythrocytes throughout the life of each twin. The identity of blood types in otherwise genetically dissimilar twins would thus result from a mixture of two kinds of erythrocytes, namely, those produced by (*a*) the twin's own cells and (*b*) cells transplanted from the co-twin. Several lines of evidence, including the *in vitro* separation of the two kinds of erythrocytes comprising the mosaic of blood cells, afford proof of this explanation. It is now found that the cellular character called J constitutes an exception to the general rule in that twins possessing a blood admixture nevertheless may differ in respect to the presence or absence of this blood factor. The present report is concerned with the reasons for this non-conformity in behavior of character J.

Each of the numerous antigenic factors of cattle erythrocytes is recognized by the reaction (hemolysis) of these cells produced by a corresponding reagent (antibodies) and complement (fresh rabbit serum). For example, erythrocytes of an individual that possesses a gene for antigen A are sensitized by the action of A reagent and subsequently destroyed (hemolysed) by the action of the complement, while those of an animal that lacks gene A are non-reactive in the presence of A reagent and complement. The blood cells of an individual are tested with each of 40-odd reagents for the presence or absence of reactivity (hemolysis). The blood type (antigenic formula) of an animal, as determined by these tests, constitutes the reactions of the cells when tested with each of the reagents. A diagrammatic representation of the techniques involved and references to previous reports are given by Stormont and Cumley.[5]

TABLE 1

BLOOD ADMIXTURE IN A PAIR OF TWINS DIFFERING IN FACTOR J

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | BLOOD TYPES | | | | | | | | | | | | | | |
| Twin no. 1: | A | $\overset{\pm}{C_1}$ | F | $\overset{\pm}{G}$ | H | J | P | $\overset{\pm}{Q}$ | $\overset{\pm}{V}$ | W | $X_2$ | Z | $\overset{\pm}{D'}$ | H′ | I′ |
| Twin no. 2: | A | $\overset{\pm}{C_1}$ | F | $\overset{\pm}{G}$ | H | | P | $\overset{\pm}{Q}$ | $\overset{\pm}{V}$ | W | $X_2$ | Z | $\overset{\pm}{D'}$ | H′ | I′ |
| | BLOOD TYPES OF THE SEPARATED CELLS FROM TWIN NO. 1 | | | | | | | | | | | | | | |
| Type 1: | A | $C_1$ | F | | H | J | P | Q | V | W | $X_2$ | Z | | H′ | I′ |
| Type 2: | A | | F | G | H | J | P | | | W | $X_2$ | Z | D′ | H′ | I′ |
| | BLOOD TYPES OF THE SEPARATED CELLS FROM TWIN NO. 2 | | | | | | | | | | | | | | |
| Type 1*a*: | A | $C_1$ | F | | H | | P | Q | V | W | $X_2$ | Z | | H′ | I′ |
| Type 2*a*: | A | | F | G | H | | P | | | W | $X_2$ | Z | D′ | H′ | I′ |

Generally, erythrocytes that are sensitized with sufficient antibody are completely hemolysed when complement is added. The cells of the twins (table 1), however, exhibited only partial hemolysis when sensitized with antibodies for the factors C, G, Q, V and D′, as indicated by the ± sign above these factors. Cell counts were made following the hemolysis of aliquots of 2.5% suspensions of the erythrocytes of each twin. These counts showed that approximately 65% of the cells of each twin were hemolysed following sensitization with either the G or D′ reagents, whereas about 35% were destroyed following sensitization with the reagents for C, Q, and V, respectively. These results demonstrated that only one kind of erythrocytes in the mixture possessed the antigenic factors G and D′ while the other kind possessed the factors C, Q and V not found in the first. The cells that were not hemolysed following sensitization with the G (or D′) reagent contained the factors C, Q and V as well as the factors common to both types of blood (i.e., A, F, H, P, W, $X_2$, Z, H′ and I′). Likewise, the other kind of cells, which were not lysed by the C reagent (or by those for Q or V), contained the factors G and D′ in addition to those common to both kinds of cells. The J character obviously did not conform to the behavior of the others, since it appeared in both kinds of the blood cells of twin no. 1 and was not demonstrable in either kind of the cells of the co-twin.

The J character is inherited as a dominant in contrast with its absence,[6] as are the other known antigenic components of the erythrocytes of cattle.

It would appear, however, that erythrocytes derived from the primordial blood cells of an individual possessing the gene for J (twin no. 1, table 1) fail to manifest this phenotype in a co-twin-host lacking the J gene. Conversely, those erythrocytes derived from an individual lacking the gene are capable of developing this character in a host possessing the gene.

The unique behavior of J in individuals of multiple births focused attention on other peculiarities of this blood factor that suggested a plausible explanation of this phenomenon.

1. The J character has been detected only by means of normal antibodies present in certain cattle that lack J.[6] Attempts to produce immune antibodies specific for J in cattle and rabbits have not been successful to date. In contrast, all the other blood factors of cattle are recognized by means of immune antibodies.

2. The J character of the erythrocytes seemingly is not present at birth, since the cells of several young calves which originally were J-negative were found in subsequent tests to be J-positive.

3. A soluble J or J-like substance has been found in considerable concentrations in the plasma of J-positive individuals. (This observation was made by Mrs. M. W. Ycas, formerly of this laboratory, who noted that the plasma of J-positive bloods was capable of inhibiting the reaction of J antibodies.) Soluble substances corresponding to the other blood-group-factors of cattle erythrocytes have not been detected.

4. Marked differences in the avidity or reactivity of the cells of different J-positive adults for the same concentrations of J antibodies have been noted. However, the degree of reactivity of the cells of a given individual appears to be relatively constant from month to month. Qualitative differences in the J substance have not been clearly demonstrable by antibody-absorption. On the other hand, the titer of J antibodies varies markedly among individuals and within the same individual from one time to another. As a consequence of such differences, weakly reactive anti-J sera do not contain sufficient antibody to cause visible hemolysis of all J-positive bloods. A similar effect may be obtained by diluting the antibodies of a more strongly reactive normal serum. Two potent normal sera (titer 1:256 and 1:512, respectively) obtained from the same cow (No. 17Z) were used for the tests described in this report. These sera were employed routinely in optimum dilution (1:4 or 1:8) for the detection of the cellular character J.

These observations suggested the explanation that J is primarily a soluble, serologically reactive substance which may be acquired by the erythrocytes of any individual (with or without the gene) on contact with blood plasma containing the soluble substance. To test this assumption, the bloods of donors lacking J, whose cells could be separated and readily distinguished from those of the recipients, were respectively transfused into selected recipients, as given in table 2.

The cells of donor 41 that were recovered from a blood sample taken from the recipient (cow no. Z21, table 2) just after the transfusion were J-negative. The cells from each of the donors (cows 41 and Z26) recovered at intervals of 2, 4, 7, 9, 14 and 21 days after transfusion of 2.5 liters of whole citrated blood had become J-positive. The cells recovered after the fourth day were indistinguishable from those of the recipients in their degree of reactivity with J antibodies.

The results suggested that the same transformation might be accomplished *in vitro*, if given sufficient time. Washed erythrocytes from J-negative cattle were suspended in freshly filtered (Seitz filter) plasma of J-positive individuals (approximately 0.4 ml. of cells to 10 ml. of plasma) and incubated at 37°C. The cells were changed to fresh plasma at intervals of 2 hours during the day and were kept at refrigerator temperature (±4° C.) without a further change overnight. The treated cells became J-positive within 24 hours. These experiments are in agreement with the proposal that the J character of cattle is primarily a soluble, serologically reactive substance that is acquired by erythrocytes in contact with this substance in the blood plasma.

TABLE 2

THE *in vivo* TRANSFORMATION OF J-NEGATIVE ERYTHROCYTES TO J-POSITIVE

| EXPERIMENT NO. 1 | BLOOD TYPES | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Donor 41 | | $C_1$ | F | G | | | $X_2$ | | $Y_2$ | Z | $E'_1$ | H′ |
| Recipient Z21 | A | $C_1$ | F | G | *J* | W | $X_2$ | | $Y_2$ | Z | $E'_1$ | H′ |
| EXPERIMENT NO. 2 | | | | | | | | | | | | |
| Donor Z26 | A | $C_1$ | F | | | $O_3$ | S | | D′ | | $E'_3$ | H′ |
| Recipient Z18 | A | $C_1$ | F | G | *J* | $O_3$ | S | Z | D′ | | $E'_3$ | H′ |
| BLOOD TYPES OF DONORS' CELLS RECOVERED 2 DAYS, AND AT LATER INTERVALS POST-TRANSFUSION | | | | | | | | | | | | |
| Donor 41 | | $C_1$ | F | G | *J* | | $X_2$ | | $Y_2$ | Z | $E'_1$ | H′ |
| Donor Z26 | A | $C_1$ | F | | *J* | $O_3$ | S | | D′ | | $E'_2$ | H′ |

Assuming that tissues which produce soluble J substance, or the cells ancestral to these tissues, are not interchanged by fetuses during embryonic life, the behavior of J in twins having mixed blood types is readily understood. Both kinds of erythrocytes in a twin possessing the gene acquire the soluble substance from the plasma, whereas both kinds of cells of the co-twin that lacks the gene are J-negative.

Holtman[7, 8] demonstrated that bacteria are capable of acquiring serologically reactive substances from culture-media and from the animal host. The acquired substances were from media containing horse serum and from guinea pig hosts. Both these species possess Forssman antigen as well as serologically reactive substances related to A of man. As Holtman pointed out, certain of the heterophile reactions attributed to bacterial cells may be consequent on the acquisition of these substances from the media or the host.

A definite serological relationship between the J substance of cattle and the A of humans was also established. Normal anti-J sera, that did not contain appreciable agglutinins for the cells of humans of groups O and B, nevertheless agglutinated, in higher titers, the cells of individuals of blood groups A and AB. Moreover, both the saliva and sera of humans of groups A and AB inhibited J antibodies in their reaction with J cells. Presumably, this was due to the presence of a soluble, serologically reactive substance in the saliva and serum corresponding to the A substance of the cells. Likewise, commercial preparations of pepsin and trypsin contained a soluble, heat-stable substance (or substances) which was capable of inhibiting J antibodies. Presumably the J substance is closely allied serologically and thereby chemically with other A or A-like polysaccharides generally distributed throughout the animal kingdom.

With few exceptions, the antigens of blood cells have behaved genetically as simple dominants in contrast to their absence. Moreover, their relative ease of detection by means of antibodies, their apparent constancy under varied environmental conditions and throughout life as well as the fact that these specific properties are contained in the stroma of erythrocytes[9] would suggest that they are produced by the action of genes in maturing erythrocytes or their cellular precursors. Indeed, the idea has been expressed[10, 11] that cellular antigens may be the more or less primary products of their causative genes. While the evidence concerning the J factor of cattle does not constitute an exception, as far as is known, to the above idea, it demonstrates that the synthesis may not always be in the cells in which a particular factor is recognized.

*Summary.*—Evidence is presented that the J factor of bovine erythrocytes is primarily a soluble, serologically reactive substance elaborated by tissues other than those which produce erythrocytes. Erythrocytes acquire the J substance on contact with soluble J in the plasma.

---

* Appreciated contributions to various phases of this study have been made by Professor M. R. Irwin. Considerable technical assistance to the project was rendered by Miss M. Claire Busk.

† From the Departments of Genetics (No. 396), Veterinary Science and Dairy Husbandry, University of Wisconsin, in coöperation with the Bureau of Animal Industry, U. S. Department of Agriculture. This study was supported in part by the Research Committee of the Graduate School from funds supplied by the Wisconsin Alumni Research Foundation. Blood samples were obtained through the generous coöperation of workers at state experiment stations and of numerous private breeders of cattle.

[1] Stormont, C., Owen, R. D., and Irwin, M. R., in preparation.

[2] Stormont, C., in preparation.

[3] Owen, R. D., *Science*, 102, 400–401 (1945).

[4] Lillie, F. R., *Ibid.*, **43**, 611–613 (1916).

[5] Stormont, C., and Cumley, R. W., *J. Hered.*, **34**, 34–41 (1943).

[6] Ferguson, L. C., Stormont, C., and Irwin, M. R., *J. Immunol.*, **44**, 147–164 (1942).

[7] Holtman, D. F., *Ibid.*, **36**, 405–412 (1939).

[8] Holtman, D. F., *Ibid.*, **36**, 413–424 (1939).

[9] Landsteiner, K., *The Specificity of Serological Reactions*, Harvard University Press, Cambridge, Mass., 1945.

[10] Irwin, M. R., and Cumley, R. W., *Am. Nat.*, **77**, 211–233 (1943).

[11] Haldane, J. B. S., *Perspectives in Biochemistry*, edited by Needham and Green, Cambridge University Press, 1938, pp. 1–11.

---

# *FURTHER STUDIES ON THE EFFECT OF INFRA-RED RADIATION ON X-RAY-INDUCED CHROMATID ABERRATIONS IN TRADESCANTIA* *

By C. P. Swanson

Department of Biology, Johns Hopkins University, Baltimore, Maryland

Communicated by K. Sax, March 21, 1949

The time limit within which x-ray-induced breaks in the chromosomes of dividing cells restitute, or recombine non-homologously, has not been established with any degree of exactitude, if, indeed, such a determination is possible. The problem has been variously attacked, but no single generalization adequately encompasses the divergent data obtained from such organisms as Tradescantia, Drosophila and maize. Even from the extensive x-ray studies which have been made on the microspore chromosomes of Tradescantia, the resultant conclusions drawn by the several investigators are only in partial agreement.

Time-intensity and fractional dosage studies on the microspore chromosomes of Tradescantia have led Sax to conclude that the broken ends remain in a condition permitting recombination or restitution for a period no longer than one hour.[1, 2] There is some indication that the time span may be as short as 20 minutes. The fact that temperature changes and centrifugation will modify the x-ray-induced frequency of aberrations in the same material *only* when used within certain time limits supports this conclusion.[3, 4] Data from experiments employing various radiations in combination provide additional supporting evidence. Ultra-violet, used as a pre-treatment, reduces the percentage of x-ray-induced aberrations in both Drosophila and Tradescantia.[5, 6] When used as a post-treatment on the pollen tube chromosomes of Tradescantia, the effectiveness of the ultra-violet in reducing the percentage of detectable aberrations diminishes as the time interval between x-ray and ultra-violet radiation increases. The double (isochromatid) deletions appeared to be unaffected by post-treatment, and hence were considered to be formed immediately upon x-radiation; the chromatid exchanges showed no reduction in number when treated with ultra-violet one hour after the breaks were induced; the single chromatid deletions continued to be reduced in number for at least an hour after their initial induction.

In addition to the direct observational evidence discussed above, Lea and Catcheside[7] and Marinelli, *et al.*,[8] have contributed mathematical support. The former have concluded that primary chromatid breaks remain in a rejoinable condition for an average period of $3^1/_2$ minutes; the latter state that "90 per cent of the single chromosome breaks would heal within 115 minutes after their occurrence." Although the time discrepancies are large, neither hypothesis seriously refutes Sax's conclusion that the maximum duration of the rejoinable period for broken ends of chromosomes is between 20 to 60 minutes after the initial rupture has occurred.

On the other hand, the findings in Drosophila and maize give definitive proof that broken chromosomes need not undergo recombination or restitution within any limited interval of time other than that imposed by the reproductive mechanism of the organism. Both Muller[9] and Kaufmann[10] have demonstrated that x-ray-induced breaks in the mature sperm cells of Drosophila do not recombine until after the sperm has entered the egg in the process of fertilization, and in Muller's work, the time lapse for some breaks must have been of a month's duration. McClintock[11] has most ingeniously provided evidence that a similar condition holds for the mechanically induced broken ends of chromosomes in the sperms and eggs of maize, since recombination between paternal and maternal chromosomes can be observed to have taken place in the zygotic nucleus. Evidence of a somewhat different, though nonetheless relevant, character can be extracted from Sparrow's[12] study of the relative sensitivity of the various chromosome stages to x-rays. His data revealed that the chromosomes of Trillium, x-rayed at meiotic metaphase I, show few or no aberrations at metaphase I or anaphase I, the latter stage having a duration of at least one day. The aberrations are observed only at the first microspore metaphase. This can only mean that recombination has been very much delayed, taking place probably in interphase or during microspore prophase, no intervening meiosis II stages being present under the experimental conditions employed. One is consequently forced to the conclusion that the time interval between breakage and recombination will vary with the type of nucleus irradiated, and that the data from maize, Drosophila, Tradescantia and Trillium cannot be equated on any single temporal basis.

The use of infra-red radiation as a supplementary tool for modifying the rate of aberrations provides a means for reinvestigating the problem in

Tradescantia, a time interval of 21 hours being imposed between x-radiation and chromosome fixation. (This time limit is arbitrarily imposed to insure the detection of chromatid aberrations *only*, and to reduce possible variance in the rate of aberrations arising out of variations in the state of the nucleus at the time of radiation.) If it is assumed that *all* of the broken ends restitute or recombine within a certain period after x-radiation, then a stage should be reached after which the application of infra-red will no longer influence the rate of observable structural changes. Preliminary studies[13] have indicated that an infra-red treatment four hours after x-radiation still effects an increase in the rate of single deletions and exchange aberrations. The present report extends these data for the full period up to the time of chromosome fixation. In addition, these data as well as other findings to be presented enable us to examine further the hypotheses

TABLE 1

FREQUENCY OF X-RAY-INDUCED CHROMATID ABERRATIONS WHEN INFRA-RED IS APPLIED AS A POST-TREATMENT. DATA IN %. 107 R

| TIME IN HRS. AFTER X-RAY | | TOTAL CHROMOSOMES | SINGLE DELETIONS | DOUBLE DELETIONS | EXCHANGES | TOTAL ABERRATIONS |
|---|---|---|---|---|---|---|
| 1 | x-ray control | 1,914 | 4.02 | 8.41 | 1.05 | 13.48 |
| | x-ray + infra-red | 1,224 | 4.57 | 10.37 | 5.39 | 20.33 |
| 2 | x-ray control | 876 | 2.74 | 10.50 | 0.91 | 14.15 |
| | x-ray + infra-red | 1,122 | 3.12 | 10.78 | 5.70 | 19.60 |
| 4 | x-ray control | 798 | 2.88 | 8.27 | 1.75 | 12.90 |
| | x-ray + infra-red | 1,080 | 3.88 | 6.57 | 6.85 | 17.20 |
| 5 | x-ray control | 1,260 | 3.08 | 4.12 | 1.90 | 9.10 |
| | x-ray + infra-red | 1,506 | 4.51 | 8.20 | 6.50 | 19.21 |
| 9 | x-ray control | 1,548 | 3.23 | 5.55 | 3.00 | 11.78 |
| | x-ray + infra-red | 1,548 | 7.00 | 7.75 | 10.85 | 25.60 |
| 12 | x-ray control | 768 | 2.38 | 4.43 | 0.26 | 7.07 |
| | x-ray + infra-red | 1,080 | 3.14 | 6.66 | 11.11 | 20.91 |
| 15 | x-ray control | 1,452 | 2.50 | 4.44 | 0.55 | 7.49 |
| | x-ray + infra-red | 1,896 | 5.12 | 5.48 | 4.80 | 15.40 |
| 18 | x-ray control | 972 | 4.32 | 9.26 | 2.47 | 16.05 |
| | x-ray + infra-red | 864 | 5.10 | 6.48 | 5.50 | 17.08 |
| Total | x-ray control | 9,588 | 3.22 ± 0.18 | 6.55 ± 0.25 | 1.52 ± 0.12 | 11.29 ± 0.32 |
| | x-ray + infra-red | 10,320 | 4.69 ± 0.21 | 7.70 ± 0.26 | 7.05 ± 0.25 | 19.44 ± 0.39 |

which have been advanced to account for the mechanism of infra-red action as it affects the behavior of x-rayed chromosomes.

*Materials and Methods.*—A clonal line of *Tradescantia paludosa* Anders. and Woodson provided the experimental material. Fixation of the chromosomes was made at 21 hours after x-radiation, regardless of the time of application of the infra-red. All x-ray treatments consisted of unfiltered radiation from an industrial type of instrument, delivered at 80 kv. and 10 ma., and at a distance of 16 inches from the center of the tungsten target. Treatment time was $1^1/_2$ minutes, and the total dose, as measured by a Victoreen dosimeter, was 107 r. The buds to be treated with infra-red were placed in a large test tube, which was then rotated slowly for three hours in a beam of infra-red emitted by a commercial drying lamp operating at 110 volts and 250 watts. The radiation, before reaching the buds, was passed through a Corning glass filter No. 2540 opaque to visible wave-lengths of light but transmitting a high percentage of the infra-red from 0.75 to 4.0 microns. 'A stream of tap water constantly flowing over the rotating tube probably absorbed all infra-red wave-lengths longer than 1.5 microns, thus limiting the effective region to that portion of the spectrum between 0.75 and 1.5 microns. To insure adequate x-ray controls within each experiment, all the buds of any particular group were x-rayed simultaneously, whether they received infra-red or not. Only metaphase figures were scored for induced aberrations, and each figure in the "Total chromosomes" columns of tables 1 and 2 represents an analysis of 3 to 7 slides from acetocarmine smear preparations.

TABLE 2

FREQUENCY OF X-RAY-INDUCED CHROMATID ABERRATIONS WHEN A DELAY EXISTS BETWEEN THE TIME OF INFRA-RED AND X-RAY TREATMENTS. DATA IN %. 107 R

| TREATMENT | TOTAL CHROMOSOMES | SINGLE DELETIONS | DOUBLE DELETIONS | EXCHANGES | TOTAL ABERRATIONS |
|---|---|---|---|---|---|
| X-ray | 1752 | 2.74 | 5.13 | 0.91 | 8.78 |
| Infra-red + x-ray: | | | | | |
| 24-hour delay | 1944 | 6.48 | 7.25 | 5.96 | 19.69 |
| 45-hour delay | 1296 | 6.93 | 8.72 | 12.04 | 27.69 |
| 73-hour delay | 1128 | 5.67 | 6.38 | 8.51 | 20.56 |
| 96-hour delay | 1284 | 6.24 | 6.67 | 7.45 | 20.36 |

*Experimental Results.*—The data on chromatid aberrations derived from the infra-red post-treatment series of previously x-rayed buds are summarized in table 1. Each time-interval treatment is given with its own x-ray control data, and, for comparative purposes, each treatment of x-rays plus infra-red should be considered in terms of its own x-ray control. The fluctuations existing between the various control treatments make such comparisons necessary.

Considering each type of aberration separately, it is evident that the greatest increases are to be found in the exchange aberrations. Except for the 18-hour series which shows a difference significant only at the 5 per cent level, the remainder of the treatments show increases which are highly significant. The increases in single and double deletions resulting from infra-red post-treatment are not as striking. The former show increases which are consistently higher throughout all treatments, while the double deletions show decreases with infra-red in the 4- and 18-hour series. The increases nevertheless are highly significant for both types of aberrations whether the *P* values are calculated by standard error methods or by summation of the individual chi squares. There can be little doubt, therefore, that infra-red post-treatments can effectively modify the rate of detectable aberrations in Tradescantia throughout the entire prophase period. Considered in the light of Sax's extensive data from time-intensity, fractional dosage and temperature experiments on identical material which definitely point to a rather brief period during which recombination can occur, these increases were somewhat unexpected.

How permanent are the chromosomal effects of infra-red which lead to a greater observable rate of x-ray-induced aberrations? The quantum energies of infra-red within the spectral limits employed are not of sufficient magnitude to produce dissociation within most molecular systems, and the induced vibrational and rotational changes of electrons are of momentary duration. It has been suggested[14] that low-grade disturbances within the polymerized nucleoproteins may "sensitize" the chromosomes to subsequent treatments with x-rays, but our lack of knowledge of the chemical organization of the chromosome, and of the possible effect of the infra-red radiation upon such a chemical system makes any inquiry as to the mechanism involved a purely speculative one. Earlier studies on Tradescantia[15] indicated, however, that the induced changes, whatever their nature might be, are at least of a semipermanent character, for an interval of 21 hours between the infra-red and x-ray treatments led to no appreciable reduction in infra-red effectiveness. The data in table 2 indicate that the time interval between the two radiations may be extended to 96 hours without reducing the infra-red effect. The changes, once induced in the chromosomes, are apparently not reversible with time. The time interval employed, however, is not of sufficient length to encompass an entire mitotic cycle. Since the induced effect can be brought about in resting nuclei and persist until prophase, the stage at which chromatid breaks are presumed to be produced by x-rays, it is possible that the "sensitization" mechanism has as its basis some alteration induced in the processes which transform the chromosomes from a resting to a prophase state. Apart from subtle chemical changes which cannot be detected cytologically, the more conspicuous phenomena are the progressive dehydration, nucleination and coiling of the chromosomes. That coiling is unaffected by the infra-red is reasonably certain; the length and diameter of the chromosomes appear the same whether infra-red has been applied or not. Little can be said, from a strictly cytological point of view, about dehydration of the chromosomes. On the other hand, if the infra-red interferes, in any way, with nucleic acid polymerization, and its attachment or detachment to the chromosome framework, a process which according to some cytologists stabilizes the chromosome during cell division, then one might theorize that any agent upsetting nucleination would tend to increase or decrease the fragility of the chromosome to x-radiation. As has been emphasized, however, whatever changes may occur in the chromosomes are not reflected in any morphologically detectable alteration of structure. The only consistent difference noted between control and infra-red-treated microspores is an increased percentage of cells observed dividing at an oblique angle. In the untreated microspores of Tradescantia, the spindle is oriented in the short axis of the cell. The presence of two large sap vacuoles, one at either end of the cell, makes this, however, the long axis of the cytoplasm. About 5 per cent of the cells normally lack one vacuole, thus producing a condition such that the long axis of the cytoplasm is oblique. The spindle orients itself accordingly. In all infra-red-treated buds, 25 per cent of the cells, on an average, possess an oblique division plane with an accompanying absence of one vacuole or both. The meaning of this observation in terms of an increased rate of chromatid aberrations is obscure, if indeed there is any correlation, unless one is dealing here with a viscocity change which affects both chromosomes and cytoplasm.

*Discussion.*—The interpretation placed upon the data presented in this report will depend upon the light in which they are viewed, i.e., whether the data are viewed in the light of the Kaufmann and Gay[16] hypothesis as to the mechanism of infra-red effect, or in the light of the hypothesis advanced by Swanson and Hollaender.[15] Consideration must also be given to the extensive data on x-rayed microspore chromosomes of Tradescantia, all of which point to the occurrence of recombination within a relatively short period of time after the initial induction of the primary breaks.

Kaufmann and Gay[16] consider that the production of primary breaks in the chromosomes of Drosophila sperm cells results *only* from the ionizing action of the x-rays. The additional infra-red radiation merely makes more of the breaks available for purposes of recombination, i.e., recombination is favored at the expense of restitution, but the infra-red does not of itself play any rôle in increasing the number of primary breaks. The principal basis upon which this hypothesis rests is the evidence that infra-red, while effectively modifying the rate of detectable recombinations, is without effect on the rates of induced dominant or recessive lethals, both presumably being single-hit phenomena. The second hypothesis,[15] which advanced the opinion that the infra-red radiation was instrumental in increasing the actual number of primary breaks, was proposed in order to relate the data for both the pre- and post-treatments to a single mechanism. It was suggested that "if both infra-red and x-rays are capable of weakening the chromosome structure (this in addition to the normally realized x-ray breaks) in a manner which would not be detectable as distinct breaks, and which would not be realized unless further disturbed by the addition of the other type of radiation, then an additional increase in break frequency would be expected with both pre- and post-treatments with infra-red radiation." Suffice it to say, no critical evidence has yet been obtained which invalidates one or the other, or both, of the hypotheses. Physical data alone would suggest that infra-red energies are incapable of disrupting ordinary molecular bonds, but our lack of knowledge as to the effects of infra-red on macromolecular systems such as the chromosome nucleoproteins makes such considerations mere speculations.

Time-intensity,[1] fractional dosage,[2] temperature[3] and centrifugation[4] data fit into the interpretation that the time interval of recombination of x-rayed microspore chromosomes of Tradescantia is comparatively brief; it is to be measured in minutes and not in hours. The data in table 1, however, definitely indicate that infra-red post-treatments effectively increase the rate of all aberrations for a time interval between the x-ray and infra-red treatments even as great as 18 hours. To account for these data on the hypothesis of Kaufmann and Gay it would have to be assumed that, after normal recombination has taken place, there still remains a reservoir of primary breaks which do not restitute and which, under the influence of infra-red, are activated to a state wherein they will enter into recombination with other similarly activated breaks or be expressed as single deletions. It must also be assumed that this reservoir of breaks cannot be tapped by other environmental agents such as temperature and centrifugation. Such an explanation is not in good agreement with the time-intensity and fractional dosage experiments.[1, 2]

On the other hand, if it be assumed that the infra-red can raise the rate of primary breaks, the data presented here and earlier[15] become explicable and at the same time remain conformable to the data of Sax. An increase in both single deletions and exchange aberrations is to be expected on this basis, whereas a decrease in single deletions would be expected to accompany an increase of exchange aberrations, if only recombination were influenced by the infra-red, since the exchanges arise from single deletion types of breaks. It is, of course, entirely possible that both hypotheses are correct. Infra-red may induce primary breaks and also affect recombination. At the moment, however, we are inclined to feel that, in light of the Tradescantia data, the essential effect of the infra-red is to induce breaks. At the same time it is recognized that this explanation is less likely when the Drosophila data are considered by themselves. Obviously more critical experiments are needed to test the two hypotheses.

It was earlier stated[15] that double deletions were unaffected by infra-red post-treatment, these being single-hit aberrations which were realized at the time of radiation with x-rays. In the light of the data presented, this view is no longer tenable.

[*] Supported by a grant from the National Advisory Health Council, United States Public Health Service.

[1] Sax, K., PROC. NAT. ACAD. SCI., **25**, 225–233 (1939).

[2] Sax, K., *Cold Spring Harbor Symposia Quant. Biol.*, **9**, 93–101 (1941).

[3] Sax, K., and Enzmann, E. V., PROC. NAT. ACAD. SCI., **25**, 397–405 (1939).

[4] Sax, K., *Ibid.*, **29**, 18–21 (1943).

[5] Kaufmann, B. P., and Hollaender, A., *Genetics*, **31**, 368–376 (1946).

[6] Swanson, C. P., *Ibid.*, **29**, 61–68 (1944).

[7] Lea, D. E., and Catcheside, D. G., *J. Genetics*, **44**, 216–245 (1942).

[8] Marinelli, L. C., Nebel, B. R., Giles, N., and Charles, D. R., *Am. J. Bot.*, **29**, 866–874 (1942).

[9] Muller, H. J., *J. Genetics*, **40**, 1–66 (1940).

[10] Kaufmann, B. P., PROC. NAT. ACAD. SCI., **27**, 18–24 (1941).

[11] McClintock, B., *Ibid.*, **28**, 458–463 (1942).

[12] Sparrow, A. H., *Ibid.*, **30**, 147–155 (1944).

[13] Swanson, C. P., *Am. J. Bot.*, **34**, 12a (1947).

[14] Kaufmann, B. P., Hollaender, A., and Gay, H., *Genetics*, **31**, 349–367 (1946).

[15] Swanson, C. P., and Hollaender, A., PROC. NAT. ACAD. SCI., **32**, 295–302 (1946).

[16] Kaufmann, B. P., and Gay, H., *Ibid.*, **33**, 366–372 (1947).

---

# THE LAST ENTRY IN GAUSS'S DIARY

BY S. CHOWLA

THE INSTITUTE FOR ADVANCED STUDY

Communicated by H. Weyl, March 24, 1949

Herglotz, in his paper "Zur letzten Eintragung im Gaussschen Tageguch" [*Leipzig. Ber.*, *73*, 271–276 (1921)] proves by means of the complex multiplication of elliptic functions, the following conjecture of Gauss (entered in his Tagebuch on July 9, 1814):

For a complex primary prime modulus $\pi \equiv 1 \bmod (1 - i)^3$ the number of solutions of the congruence (including the solutions $x, y \equiv \pm i, \infty;\ \infty, \pm i$):

$$x^2y^2 + x^2 + y^2 \equiv 1 \pmod{\pi}$$

is equal to the norm of $\pi - 1$. Dedekind verified this for $N(\pi) < 100$. We give a proof (Theorem 1 below) based on elementary number theory. NOTATION: $p$ denotes a prime $\equiv 1 \pmod 4$; $g$ is a primitive root of $p$;

$(n/p)$ is the Legendre symbol of quadratic residuacity; $\theta$ is defined uniquely by $p = \theta^2 + \phi^2$, $\theta \equiv 1$ or $3 \pmod 4$ according as $p \equiv 5$ or $1 \pmod 8$; finally

$$S_m = \sum_{x=1}^{p-1} \left(\frac{x^4 - g^m}{p}\right) \qquad [m = 0, 1, 2, 3].$$

THEOREM 1:

$$S_0 = 2\theta - 2.$$

*Proof:* We observe that our theorem follows immediately from the 5 relations:

(A) $S_1 = -S_3$

(B) $S_0 + S_2 = -4$

(C) $S_0^2 + S_1^2 + S_2^2 + S_3^2 = 8p + 8$

(D) $S_0 \equiv 0, 4 \pmod 8$ $[p \equiv 5, 1 \pmod 8)]$

(E) $p = \theta^2 + \phi^2$; $\theta \equiv 1, 3 \pmod 4$ $[p \equiv 5, 1 \pmod 8)]$

(A) $$\sum_{1}^{p-1} \left(\frac{x^4 - g}{p}\right) = \sum_{1}^{p-1} \left(\frac{1 - gy^4}{p}\right) = -\sum_{1}^{p-1} \left(\frac{g^3 - g^4y^4}{p}\right) = -\sum_{1}^{p-1} \left(\frac{g^3 - t^4}{p}\right)\left[x = \frac{1}{y} = \frac{g}{t}\right]$$

(B) $$\frac{p-1}{4}(S_0 + S_1 + S_2 + S_3) = \sum_{a=1}^{p-1}\sum_{x=1}^{p-1}\left(\frac{x^4 - a}{p}\right) = \sum_{x=1}^{p-1}\sum_{a=1}^{p-1}\left(\frac{x^4 - a}{p}\right) = -(p-1)$$

(C) $$\frac{p-1}{4}(S_0^2 + S_1^2 + S_2^2) + S_3^2 = \sum_{a=1}^{p-1}\sum_{x=1}^{p-1}\sum_{y=1}^{p-1}\left(\frac{x^4 - a}{p}\right)\left(\frac{y^4 - a}{p}\right) = \sum_{x=1}^{p-1}\left\{\sum_{y=1}^{p-1}\sum_{a=1}^{p-1}\left(\frac{a^2 - a(x^4 + y^4) + x^4y^4}{p}\right)\right\}$$

Consider the sum inside the braces. If $y^4 \equiv x^4$, the innermost sum $= p - 2$; for the remaining $(p - 5)$ values of $y$ the innermost sum is $-2$ since writing $f = {}^1/_2(x^4 - y^4)$ this sum easily transforms into

$$-1 + \sum_{a=1}^{p} \left(\frac{a^2 - f^2}{p}\right) = -1 + \sum_{t=1}^{p-1}\left(\frac{t}{p}\right)\left(\frac{t+2}{p}\right) = -1 + \sum_{t=1}^{p-1}\left(\frac{2l+1}{p}\right) = -2[a = t + 1;\ tl = 1;\ f \not\equiv 0].$$

(D) This follows at once by noting that $x^4$ has the same non-zero value mod $p$ for exactly 4 incongruent values of $x$, and these $x$ are equidistant from ${}^1/_2p$; further exactly 4 terms of $S_0$ vanish, while the rest are $+1$ or $-1$.

Similarly we prove:

THEOREM 2:

$$\sum_{x=1}^{p-1}\left(\frac{x^6-1}{p}\right) = -2-4\theta$$

where $\theta$ is defined uniquely by $p = \theta^2 + 3\phi^2$, $\theta \equiv 1 \pmod 3$ and $p$ is a prime $\equiv 1 \pmod{12}$.

I should like to thank Professor Siegel (who found slightly more general results) and Professor Weyl for their interest in this note.

*Note added in proof:* Results similar to Theorems 1 and 2 have been found recently by E. Lehmer and A. L. Whiteman. Deuring, in a paper dedicated to Herglotz, finds more general results by advanced methods [see *Hamburg Abhandlungen*, **14**, 1941, 197–272].

---

# ON THE THEOREM OF JENTZSCH

By Aryeh Dvoretzky

Hebrew University of Jerusalem and Institute for Advanced Study, Princeton, New Jersey

Communicated by J. von Neumann, March 9, 1949

*1.* Let

$$f(z) = \sum_{n=0}^{\infty} a_n z^n, \qquad \limsup_{n\to\infty} |a_n|^{1/n} = 1, \tag{1}$$

be a power series with unit radius of convergence. Denote by

$$S_n(z) = \sum_{\nu=0}^{n} a_\nu z^\nu \qquad (n = 0, 1, 2, \ldots) \tag{2}$$

the partial sums of (1) and by $\Delta_n$ the distance of $z = 1$ from the set of roots of $S_n(z) = 0$. The classical result of R. Jentzsch[1] that every point of the circle of convergence is a limit point of zeros of partial sums may be written in the form:

$$\liminf_{n=\infty} \Delta_n = 0. \tag{3}$$

*2.* It is *impossible* to improve on (3) by exhibiting a sequence $(\varphi_n)$ of positive numbers tending (however slowly) to infinity and such that

$$\liminf_{n=\infty} \varphi_n \Delta_n = 0$$

for *all* series (1). It is, however, possible to give explicit estimates of the rate with which (a suitable subsequence of) $\Delta_n$ approaches zero in terms of the moduli of the coefficients of (1) and of its behavior near $z = 1$. This is an immediate corollary of more precise results, some of which are given below.

*3.* A function $f(z)$ regular in $|z| < 1$ (or in its intersection with a neighborhood of $z = 1$) is said to satisfy Condition C if

$$\log f(z) = o\left(\log \frac{1}{|1 - z|}\right)$$

as $z \to 1$ through some angle with vertice at $z = 1$ and one of whose sides contains the negative axis.[2]

Condition C is clearly satisfied if $f(z)$ is bounded and bounded away from zero as $z \to 1$ through the above angle, thus in particular if $f(z)$ is regular at $z = 1$ and $f(1) \neq 0$ or if $\sum a_n$ is convergent (or Cesaro summable of any order) to a sum different from zero.

*4.* First we give a result pertaining to the special case when the largest coefficients $a_n$ are comparable to some power of $n$.

THEOREM 1. *Let $f(z)$ satisfy Condition C and, furthermore, let the coefficients of its Taylor series (1) satisfy*

$$-\infty < p = \limsup_{n \to \infty} \frac{\log |a_n|}{\log n} < \infty. \tag{4}$$

*Then there exist infinitely many $n$ for which*

$$S_n\left(1 + \frac{\zeta}{n} \log n\right) = 0 \tag{5}$$

*has roots in the rectangle*

$$-p - 1 - \delta < \xi = Re\{\zeta\} < -p + \delta, \qquad -\delta < \eta = Im\{\zeta\} < \delta, \tag{6}$$

*$\delta$ being an arbitrary positive number.*

*Moreover, if $(n_\nu)_{\nu=1}^{\infty}$ is a sequence of increasing integers for which*

$$\lim_{\nu = \infty} \frac{\log |a_{n_\nu}|}{\log n_\nu} = p, \tag{7}$$

*then, for all $\nu$ sufficiently large, (5) for at least one of $n = n_\nu$ and $n = n_\nu - 1$ has roots in (6).*

Thus we know of the existence of a subsequence of $(\Delta_n)$ for which the order of magnitude of $\Delta_n$ does not exceed $\frac{\log n}{n}$. Simple examples, e.g., $f(z) = (1 + z)^{-2}$, show this to be the correct order. Moreover, the location of some roots is known to within $\left(\frac{1}{2} + o(1)\right)\frac{\log n}{n}$ which, unlike the estimate of $\Delta_n$, is independent of the value of $p$.

*5.* In the general case, when (6) is not assumed, there is the following analogous result.

THEOREM 2. *Let (1) satisfy Condition C then there exist infinitely many $n$ (with $a_n \neq 0$) for which at least one of the two equations*

$$S_n\left(|a_n|^{-1/n} + \frac{\zeta}{n}\log n\right) = 0, \qquad S_{n-1}\left(|a_n|^{-1/n} + \frac{\zeta}{n}\log n\right) = 0 \quad (8)$$

*has roots in the rectangle*

$$-1 - \delta < \xi = Re\{\zeta\} < \delta, \qquad -\delta < \eta = Im\{\zeta\} < \delta, \quad (9)$$

*$\delta$ being an arbitrary positive number.*

Disregarding possible roots at $z = 0$, $|a_n|^{-1/n}$ is approximately the geometric mean of the moduli of the roots of $S_n(z) = 0$. It is remarkable that through consideration of distances from this geometric mean, instead of from $z = 1$, we have recovered in the general case the order $\frac{\log n}{n}$ encountered in *4*. This, however, does not yet yield any information, of the form considered in *2*, about the $\Delta_n$. Indeed, as far as the statement of theorem 2 goes, it is conceivable that the values of $n$ the existence of which is asserted are such that lim sup $|a_n|^{1/n} < 1$, and thus not even (3) is implied.

*6*. To overcome this failing it is necessary to specify further the values of $n$ for which the assertion of theorem 2 is made. It turns out that these are those $n$ for which the $a_n$ are "large." Some sufficient conditions for this "largeness" are listed below:

An infinite sequence $(n_\nu)_{\nu=1}^{\infty}$ of increasing integers is an $\mathcal{L}^1$ $(a_n)$ sequence if it satisfies[3]

$$\max_{j \leq n_\nu}\left(\log|a_j| - \frac{j}{n_\nu}\log|a_{n_\nu}|\right) = O(\log n_\nu),$$

and

$$\sup_{j \geq n_\nu}\frac{1}{j}\log|ja_j| = \frac{1}{n_\nu}\log|n_\nu a_{n_\nu}| + o\left(\frac{\log n_\nu}{n_\nu}\right).$$

$(n_\nu)$ is an $\mathcal{L}^2(a_n)$ sequence if to every $\epsilon > 0$ there corresponds a $\gamma > 0$ such that

$$\sup_{j \geq n_\nu}\left(\log|a_j| - \frac{j}{n_\nu}\log|a_{n_\nu}| + \gamma\frac{j}{n_\nu}\log n_\nu\right) \leq (\epsilon + o(1))\log \eta_\nu.$$

$(n_\nu)$ is an $\mathcal{L}^3(a_n)$ sequence if there exist an integer $N$ and a positive $\gamma$ such that

$$\max_{N \leq j \leq n_\nu}\left(\frac{1}{j}\log|a_j| + \frac{1+\gamma}{j}\log j\right) = \frac{1}{n_\nu}\log|a_{n_\nu}| + \frac{1+\gamma+o(1)}{n_\nu}\log n_\nu.$$

Finally, an infinite sequence of increasing integers is an $\mathcal{L}(a_n)$ sequence if it is the union of three sequences each of which is either empty or an $\mathfrak{L}^k(a_n)$ sequence ($k = 1, 2, 3$).

It can be shown that whatever the series (1) there exist $\mathfrak{L}(a_n)$ sequences. It is also easily seen that if $(n_\nu)$ is an $\mathfrak{L}(a_n)$ sequence then

$$\lim_{\nu = \infty} |a_{n_\nu}|^{1/n_\nu} = 1. \tag{10}$$

Furthermore it is immediately verified that when (4) holds and $(n_\nu)$ satisfies (7) then it is an $\mathfrak{L}(a_n)$ sequence.

*7.* With this concept we have corresponding to (and including) the second part of theorem 1

THEOREM 3. *Let the conditions of theorem 2 be satisfied and $(n_\nu)$ be an $\mathfrak{L}(a_n)$ sequence, then for all $n = n_\nu$ with large $\nu$ at least one of the equations (8) has roots in (9).*

By choosing $\mathfrak{L}(a_n)$ sequences for which the approach of (10) to the limit is as fast as possible, explicit estimates of the type discussed in *2* of $\Delta_n$ are obtained.

*8.* Condition C is the only restriction on (1) in theorems 2 and 3. If it is not satisfied but there exists a real $q$ for which $(z - 1)^q f(z)$ satisfies Condition C the effect is a shifting of the rectangle (9) by $\frac{q}{n} \log n$. However, there are of course functions (1) which do not satisfy any restrictions of this type. Our methods yield, however, explicit estimates in the most general cases, we quote only the following result.

THEOREM 4. *Let (1) satisfy*

$$-\infty \leq \limsup_{n = \infty} \frac{\log |a_n|}{\log n} < -1, \tag{11}$$

*let $\beta > 1$ be arbitrary and $(\beta_n)_1^\infty$ be such that $\beta_n \geq \frac{-1}{n} \log |a_n|$ and $\left(\beta_n - \frac{\beta}{n} \log n\right)_1^\infty$ is a downward convex sequence of positive numbers. Let $0 < \vartheta < 1$ be arbitrary and denote by $M_n$ the maximum of $f(z)$ in $|z| \leq 1 - \vartheta\beta_n$, $|z - 1| < \beta_n$, and let*

$$\liminf_{n = \infty} \frac{1}{n\beta_n} \log M_n \leq 0.$$

*Then*

$$\liminf_{n = \infty} \frac{\Delta_n}{\beta_n} \leq 1. \tag{12}$$

Whenever (11) holds, sequences $(\beta_n)$ satisfying all the above requirements can be found. Since $n\beta_n \to \infty$, (12) is an improvement of (3) of the type mentioned in *2*.

*9.* The previous results can be extended in yet another direction illustrated by

THEOREM 5. *Let (1) satisfy (4) with* $p < -1$ *and* $(r_n)_1^\infty$ *be any sequence of positive numbers subject to the only restriction* $\log r_n = o(\log n)$ *as* $n \to \infty$. *Then there exist infinitely many n for which* $w = S_n\left(1 + \frac{\zeta}{n}\log n\right)$ *maps (6) on a domain* $W_n$ *over the w-plane completely covering the ring*

$$\frac{1}{r_n} < \left|w - \sum_{n=0}^{\infty} a_n\right| < r_n, \tag{13}$$

*and, indeed, more than* $\sigma \log n$ *times,* $\sigma$ *being a suitable positive number.*

If instead of (6) larger domains are considered, (13) can also be replaced by larger domains. If, ultimately, (6) is replaced by $|z - 1| < \epsilon$ then (13) may be replaced[4] by $|w| < R_n$ where the $R_n$ satisfy $\log R_n = o(n)$ as $n \to \infty$ and also $\sigma \log n$ by $\sigma n$.

*10.* When the coefficients of (1) are such that (for $\mathfrak{L}(a_n)$ sequences) the approach to the limit in (10) is slow, e.g., $a_n = e^{\sqrt{n}}$, the rate of $\Delta_n \to 0$ is also slow, however the size of (6) can be diminished. As an example of yet another phenomenon we bring the following result on gap series.

THEOREM 6. *Let* $(\lambda_n)_1^\infty$ *be a sequence of increasing integers satisfying* $\log(\lambda_n/(\lambda_n - \lambda_{n-1})) = o(\log n)$ *as* $n \to \infty$. *Let* $f(z) = \sum a_n z^{\lambda_n}$ *satisfy*[5]

$$-\infty < \bar{p} = \limsup_{n\to\infty} \frac{\log |a_n|}{\log n} < -1$$

*and* $\sum_{n=0}^{\infty} a_n \neq 0$. *Then there are infinitely many n for which*

$$S_{\lambda_n}\left(1 + \frac{\zeta}{\lambda_n}\log n\right) = 0$$

*has roots in the circle*

$$\left|\zeta - \frac{\bar{p}}{\lambda_n}\log n\right| < \frac{\delta}{\lambda_n}\log n, \tag{14}$$

$\delta$ *being an arbitrary positive number.*

The main improvement over theorem 1 is the appearance of $\log n$, instead of $\log \lambda_n$, in (14).

*11.* Condition C can be considerably weakened without impairing the order $\frac{\log n}{n}$, on the other hand if stronger assumptions are made this order can be improved; ultimately, when regularity or certain convergence type conditions are imposed, even to $\frac{1}{n}$. E.g.,

THEOREM 7. *Let (1) be regular and different from zero at* $z = 1$, *and its coefficients satisfy*

$$0 < A = \limsup_{n = \infty} |a_n| < \infty. \tag{15}$$

*Then*

$$\liminf_{n = \infty} n\Delta_n < B < \infty \tag{16}$$

*Moreover, B may be chosen as a function of the single argument* $|f(1)|d/A$ *where d denotes the distance from* $z = 1$ *to the nearest singularity of* $f(z)$.[6]

Similarly when a convergence type condition replaces regularity we have

THEOREM 8. *Let (1) satisfy (15) and* $f(z)$ *be bounded and bounded away from zero as* $z \to 1$ *through some angle with vertice at* $z = 1$ *and containing points in* $|z| < 1$. *Then if the sequence* $S_n(1)$ *is bounded, (16) again holds.*

Now, of course, instead of dependence on $f(1)$ and $d$ which have no meaning in this case, $B$ will depend on the bound of $S_n(1)$.

*12.* The shift mentioned at the beginning of *8* in the location of the roots of the partial sums when Condition C fails, is most clearly exhibited in the following theorem.

THEOREM 9. *Let (1) satisfy the conditions of theorem 7 except that* $z = 1$ *be a zero of order m of* $f(z)$. *Then*

$$\liminf_{n = \infty} n\Delta_n \left(1 - m\frac{\log n - \log\log n}{n}\right) < \infty, \tag{17}$$

*where* $\Delta_n(z_n)$ *denotes the distance from* $z_n$ *to the set of roots of* $S_n(z) = 0$.

In particular we have in this case for the order of magnitude of (a suitable subsequence of) $\Delta_n$ the upper bound $\frac{\log n}{n}$. That all roots of the partial sums may indeed be as far from $z = 1$ as allowed by (17), is seen from examples such as $f(z) = \frac{1}{1 + z} - \frac{1}{2}$.

*13.* The main tools used in establishing the above results are some of the simpler propositions of the theory of normal families or related inequalities of the theory of functions of a complex variable. Due to the general applicability of these tools, our results can be extended to much wider classes of developments than power series. In particular they can be carried over to general Dirichlet series (cf. reference 4) or even to Laplace-Stieltjes transforms. Also, instead of studying the partial sums of (1), it is possible to study in the same way the remainders or sums of blocks of terms from (1), etc., and similarly for more general developments.

Further results and proofs will be published elsewhere.

---

[1] *Acta Math.*, **41**, 219–251 (1918).

[2] The assumption that the angle contains points interior to the unit circle is sufficient; the stronger assumption in the text is made for normalization purposes, it affects the constants involved but not the order of magnitude.

[3] Unless otherwise stated, $O$ and $o$ refer to $\nu \to \infty$.

[4] Dvoretzky, A., *Studies on General Dirichlet Series* (Hebrew and English abstract), Jerusalem, 1941.

[5] This is, of course, quite different from (4).

[6] Only singularities of a principal branch need be considered.

---

# *ON DEVIATIONS BETWEEN THEORETICAL AND EMPIRICAL DISTRIBUTIONS**

BY M. KAC

DEPARTMENT OF MATHEMATICS, CORNELL UNIVERSITY

Communicated by H. Weyl, March 18, 1949

1. The determination of a theoretical distribution function $\sigma(u)$ by drawing a random sample is a familiar technique. The reliability of this procedure is based on considerations of the least upper bound of the absolute value of the difference between $\sigma(u)$ and the empirical distribution $\sigma_n(u)$, defined as $1/n$ times the number of sample elements falling below $u$. The theory requires delicate and powerful mathematical tools. Considerable simplification can, however, be achieved by taking samples whose size is not fixed but drawn from a Poisson population. This note is devoted mainly to the exposition of this idea. We also consider different measures of deviations and compare the theories for samples of fixed and random sizes.

2. Let $X_1, X_2, \ldots$ be independent random variables having the same continuous distribution functions $\sigma(u)$. Let furthermore $\psi_u(x)$ be 0 or 1 according as $x > u$ or $x < u$. The empirical distribution function $\sigma_n(u; X_1, \ldots, X_n) \equiv \sigma_n(u)$ is defined by the formula

$$\sigma_n(u) = \frac{1}{n} \sum_{j=1}^{n} \psi_u(X_j). \tag{1}$$

Kolmogoroff[1] has shown that

$$\lim_{n \to \infty} \text{Prob.}\left\{ \underset{-\infty<u<\infty}{\text{l.u.b}} |\sigma_n(u) - \sigma(u)| < \frac{\alpha}{\sqrt{n}} \right\} = P(\alpha), \tag{2}$$

where

$$P(\alpha) = \frac{\sqrt{2\pi}}{\alpha} \sum_{k=0}^{\infty} e^{-(2k+1)^2\pi^2/8\alpha^2}. \tag{3}$$

An elementary proof of this and some related results was subsequently given by Feller.[2] Quite recently Doob[3] derived Kolmogoroff's result using the theory of stochastic processes. He pointed out that

$$P(\alpha) = \text{Prob.} \left\{ \underset{0 \leq u \leq 1}{\text{Max.}} |y(u)| < \alpha \right\}, \tag{4}$$

where $y(u)$ $(0 \leq u \leq 1)$ is a Gaussian process satisfying the conditions,

$$y(0) = y(1) = 0, E\{y(u)y(v)\} = \text{Min.}\ (u, v) - uv. \tag{5}$$

Since the process $y(u)$ is related in a simple way to the classical Wiener process the calculation of (4) can be reduced to a more or less standard calculation in the diffusion theory. Doob's justification of (4) is heuristic and he states that a rigorous proof would be quite delicate.

Independently of Doob, the present author also used the reduction to the process $y(u)$ in connection with a related problem (see § 4 of this note).

3. Let $N$ be a random variable having Poisson's distribution with mean $\mu$ and independent of the collection $X_1, X_2, \ldots.$ We define the modified empirical distribution function $\sigma_\mu^*(u)$ by the formula

$$\sigma_\mu^*(u) = \frac{1}{\mu} \sum_{j=1}^{N} \psi_u(X_j), \tag{6}$$

with the understanding that if $N = 0$ the sum is to be taken as 0. The result which corresponds to that of Kolmogoroff is

$$\lim_{\mu \to \infty} \text{Prob.} \left\{ \underset{-\infty < u < \infty}{\text{l.u.b.}} |\sigma_\mu^*(u) - \sigma(u)| < \frac{\alpha}{\sqrt{\mu}} \right\} = P^*(\alpha), \tag{7}$$

where

$$P^*(\alpha) = \text{Prob.} \left\{ \underset{0 \leq u \leq 1}{\text{Max.}} |x(u)| < \alpha \right\} = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} e^{-(2k+1)^2\pi^2/8\alpha^2} \tag{8}$$

and $x(u)$ denotes the Wiener process $(x(0) = 0)$. An intuitive justification of (8) is quite simple. We first note, following Kolmogoroff, that as long as $\sigma(u)$ is continuous, the probability on the left side of (7) is independent of $\sigma(u)$. Thus it is sufficient to restrict oneself to the case when $\sigma(u)$ is the uniform distribution, i.e., $\sigma(u) = u$, $0 \leq u \leq 1$.

Next, it is seen that the process

$$x_\mu(u) = \sqrt{\mu} \left\{ \frac{1}{\mu} \sum_{1}^{N} \psi_u(X_j) - u \right\}, \qquad 0 \leq u \leq 1 \tag{9}$$

is a process with *independent increments*[4] and

$$E\{x_\mu(u)\} = 0, \qquad E\{x^2_\mu(u)\} = u. \tag{10}$$

Moreover, as $\mu \to \infty$, the probability distribution of $x_\mu(u)$ becomes normal[5] with mean 0 and variance $u$. Thus it is natural to expect that the limiting properties of the process $x_\mu(u)$ will be those of a Gaussian process

$x(u)$ $(x(0) = 0)$ with *independent increments* and such that $E\{x(u)\} = 0$, $E\{x^2(u)\} = u$. These properties are characteristic of a Wiener process.

To justify (8) rigorously one needs to prove that

$$\lim_{\mu \to \infty} \lim_{\mu \to \infty} \text{Prob.}\left\{ \underset{1 \leq k \leq 2r}{\text{Max.}} \left|x_\mu\left(\frac{k}{2^r}\right)\right| < \alpha \right\} = \text{Prob.}\ \{ \underset{0 \leq u \leq 1}{\text{Max.}} |x(u)| < \alpha\} \tag{11}$$

whereas the corresponding fact with the limiting processes interchanged (first $\mu \to \infty$ then $r \to \infty$) is trivial. Due to the fact that $x_\mu(u)$ is a process with *independent increments* (11) can be proved quite easily. In fact, one need only apply, with minor modifications, the reasoning given on page 295 of the note of Erdös and the present author, "On Certain Limit Theorems of the Theory of Probability."[7] The analogue of (11) for the process $y(u)$ is difficult and it is this step which is left out from Doob's considerations.

The deviation between the empirical and the theoretical distributions can be studied in various other ways. The following seems to be of some independent interest.

Let $V_a(x) = 1 - \psi_a(x)$ and consider the random variable

$$K_a = \int_{-\infty}^{\infty} V_a(\sqrt{\mu}(\sigma_\mu^*(u) - \sigma(u))d\sigma(u) = \int_0^1 V_a(x_\mu(u))du. \tag{12}$$

This random variable has a simple interpretation. The graph of $\sigma_\mu^*(u)$ is a step function and it is convenient to consider the jumps as vertical parts of the graph. $K_a$ represents then the sum of lengths of those pieces of the vertical parts of the graph which lie above $\sigma(u) + a/\sqrt{\mu}$.

It is easy to prove that

$$\lim_{\mu \to \infty} \text{Prob.}\ \{K_a < \alpha\} = \text{Prob.}\ \{\textstyle\int_0^1 V_a(x(u))du < \alpha\}. \tag{13}$$

A general theory of calculating probability distributions of Wiener functionals of the type $\int_0^1 V(x(u))du$ has been recently given by the present author.[6] For $a = 0$ one gets a particularly simple result (*loc. cit.*,[8] Example 2)

$$\lim_{\mu \to \infty} \text{Prob.}\ \{K_0 < \alpha\} = \frac{2}{\pi} \text{arc sin } \sqrt{\alpha}. \tag{14}$$

4. Results analogous to (14) can be also obtained for empirical distributions based on samples of fixed size. In this case

$$K_a = \int_{-\infty}^{\infty} V_a(\sqrt{n}(\sigma_n(u) - \sigma(u))d\sigma(u)$$

has the same interpretation as above, but

$$\lim_{n \to \infty} \text{Prob. } \{K_n < \alpha\} = \text{Prob. } \{ \textstyle\int_0^1 V_a(y(u))du < \alpha \}, \tag{15}$$

where $y(u)$ is the Gaussian process defined by (5). Unlike in the case of the least upper bound the justification of (15) is not difficult. The calculations of the probability distributions of $\int_0^1 V_a(y(u))du$ can be reduced to considerations of the paper (*loc. cit.*[8]) but owing to the boundedness of $V_a$ the theory can be simplified.

Let

$$Q_0(x, t) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t}$$

$$Q_{n+1}(x, t) = \int_0^t \int_{-\infty}^{\infty} \frac{e^{-(x-\xi)^2/2(t-\tau)}}{\sqrt{2\pi(t-\tau)}} V_a(\xi) Q_n(\xi, t) d\xi d\tau. \tag{16}$$

It is easy to verify that

$$E\{(\textstyle\int_0^1 V_a(y(u))du)^n\} = \sqrt{2\pi} n! Q_n(0, 1). \tag{17}$$

Let furthermore

$$Q(x, t) = \sum_{n=0}^{\infty} (-1)^n v^n Q_n(x, t) \tag{18}$$

and note that whenever $|V(\xi)| \leq M$ (in our case $M = 1$) one has

$$0 \leq Q_n(x, t) \leq \frac{(Mt)^n}{n!} Q_0(x, t). \tag{19}$$

Thus $Q(x, t)$ is well defined and

$$0 \leq Q(x, t) \leq e^{Mt} Q_0(x, t).$$

Moreover,

$$\sqrt{2\pi}\, Q(0, 1) = E\{e^{-v\int_0^1 V_a(y(u))du}\}. \tag{20}$$

From (16) and (18) it follows that

$$Q(x, t) + v\int_0^t \int_{-\infty}^{\infty} \frac{e^{-(x-\xi)^2/2(t-\tau)}}{\sqrt{2\pi(t-\tau)}} V_a(\xi) Q(\xi, \tau) d\xi d\tau = \frac{e^{-x^2/2t}}{\sqrt{2\pi t}} \tag{21}$$

and from (19) that the Laplace transform

$$\varphi(x) = \int_0^{\infty} e^{-st} Q(x, t)dt, \qquad s > M,$$

exists.

Thus taking Laplace transforms of both sides of (21) we obtain

$$\varphi(x) + \frac{v}{\sqrt{2s}} \int_{-\infty}^{\infty} \varphi(\xi) V_a(\xi) e^{-\sqrt{2s}|x-\xi|} d\xi = \frac{e^{-\sqrt{2s}|x|}}{\sqrt{2s}} \tag{22}$$

The *unique* solution of (22) can be shown to satisfy the differential equation[9]

$$^1/_2\varphi''(x) - (s + vV_a(x))\varphi(x) = 0, \qquad x \neq 0, \qquad x \neq a, \quad (23)$$

and the conditions

$$\lim_{x \to \pm\infty} \varphi(x) = 0, \qquad |\varphi'(x)| < A, \qquad \varphi'(-0) - \varphi(+0) = 2.$$

Moreover, $\varphi'(x)$ should be continuous for all $x \neq 0$.

In the simple $a = 0$ it follows that

$$\varphi(x) = \frac{\sqrt{2}}{\sqrt{s+v} + \sqrt{s}}\, e^{-\sqrt{2(s+v)}x}, \qquad x > 0$$

$$\varphi(x) = \frac{\sqrt{2}}{\sqrt{s+v} + \sqrt{s}}\, e^{\sqrt{2s}x}, \qquad x < 0,$$

and therefore

$$\varphi(0) = \int_0^\infty e^{-st}Q(0, t)dt = \frac{\sqrt{2}}{\sqrt{s+v} + \sqrt{s}} = \frac{\sqrt{2}}{v} \sum_{k=1}^\infty \binom{^1/_2}{k} \frac{v^k}{s^{k-{}^1/_2}}.$$

thus

$$Q(0, t) = \frac{\sqrt{2}}{v} \sum_{k=1}^\infty \binom{^1/_2}{k} \frac{t^{k-{}^1/_2}}{\Gamma(k - {}^1/_2)}\, v^k$$

and in particular,

$$Q(0, 1) = \frac{\sqrt{2}}{2\Gamma(^1/_2)}\, \frac{1 - e^{-v}}{v}$$

It finally follows from (20) that $\int_0^1 V_0(y(u))du$ is uniformly distributed in (0, 1).

---

* Work done under an ONR contract.

[1] "Sulla determinazione empirica di una legge di distribuzione," *Inst. Ital. Attuari, Giorn.*, **4**, 1–11 (1933).

[2] "On the Kolmogorov-Smirnov Limit Theorems for Empirical Distributions," *Ann. Math. Stat.*, **19**, 177–189 (1948).

[3] To appear in the *Ann. Math. Stat.*

[4] Hurwitz, H., Jr., and Kac, M., "Statistical Analysis of Certain Types of Random Functions," *Ann. Math. Stat.*, **15**, 173–181 (1944). It is shown in this paper that $N$ having Poisson's distribution is not only sufficient but also necessary for $x_\mu(u)$ to have independent increments.

[5] *Loc. cit.*,[4] pp. 176–177. More general results in this direction were recently published by H. Robbins, "The Asymptotic Distribution of the Sum of Random Number of Random Variables," *Bull. Amer. Math. Soc.*, **54**, 1151–1161 (1948).

[6] *Loc. cit.*,[4] p. 177.

[7] *Bull. Am. Math. Soc.*, **52**, 292–302 (1946).

[8] "On the distributions of certain Wiener functionals," *Trans. Am. Math. Soc.*, **65**, 1-13 (1949).

[9] It is this equation which, in a different way and under more general conditions, is derived, *loc. cit.*[8]

---

# *ON INTERPOSITION AS A CUE FOR THE PERCEPTION OF DISTANCE**

By P. Ratoosh

Psychological Laboratory, Columbia University

Communicated by C. H. Graham, March 26, 1949

Interposition is usually defined[1-3] as the cue for the perception of the relative distances of two objects that results when one object partially obscures or overlaps the outline of another object. Such a definition is not satisfying, for it appeals to an unknown factor: the basis for discriminating between the object that obscures and the object that is obscured.

Brown[4] describes the cue in terms of completeness of outline: "We therefore perceive objects which are complete in outline as placed in front of similar objects which are incomplete." If it is added that knowledge of completeness of outline is a function of past experience, then Brown's specification of interposition becomes valid, albeit not fruitful. However, it probably does not apply in cases where there is no familiarity with the objects outlined; yet cues for relative position may remain in these cases.

Helmholtz,[5] in 1866, pointed out that even with objects whose forms are unfamiliar, "the mere fact that the contour line of the covering object does not change its direction where it joins the contour of the one behind it, will generally enable us to decide which is which." Helmholtz's specification is given in terms of essential stimulus characteristics, and it is the writer's belief that the statement specifies the cue in all cases, not merely when the objects are unfamiliar. The purpose of the present account is to state Helmholtz's description precisely and to point out some of its consequences.

Consider a projection of figures onto a plane perpendicular to the visual axis such that the nodal point of the eye is the point of projection. In such a perspective transformation all objects will be represented by simple closed curves. (Without any loss in generality only two figures will be treated here; the same reasoning may be extended to apply to any number of objects.)

A necessary condition for interposition to become effective as a cue is that the curves of the two objects in the plane of projection have a common boundary. A point at which the boundaries of the curves meet will be called a point of intersection. Clearly there will be two points of intersection for a common boundary.

Helmholtz's specification implies that the behavior of the functions represented by the closed curves at the points of intersection alone determines which object will be seen as nearer. More specifically Helmholtz's assertion may be taken to indicate that continuity of the first derivative of the object's contour at the points of intersection is the sole determiner of relative distance. (A first derivative $f'(x)$ of a function $f(x)$ is continuous at a point $p$ if and only if for any $\epsilon > 0$, there exists a $\delta$ such that $|f'(x) - f'(p)| < \epsilon$ in the interval $|x - p| < \delta$.)[6]

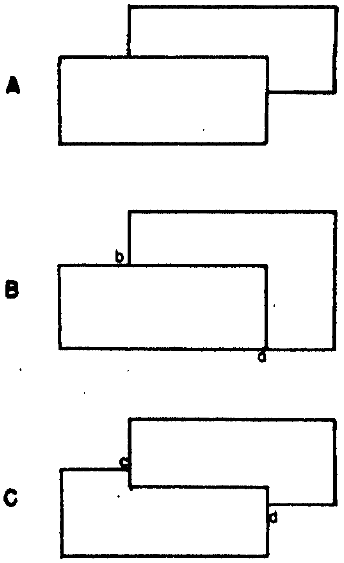


FIGURE 1

Examples of the operation of interposition as a cue for the perception of relative distance. *A* and *B* give obvious cues. The cues in *C* result in an ambiguity.

This formulation predicts that what happens at one point of intersection is independent of what happens at the other. That is to say, both points of intersection may give the same cue, both may offer no cue, only one may give a cue, or each may provide a cue which contradicts the other. If, at a point of intersection, the contours of both objects are the same with respect to continuity of their first derivatives, no cue will be provided at that point of intersection; if this occurs at both points of intersection, interposition will afford no cue for the relative distances of the two objects.

The cases in which interposition provides a cue may be classified in the following manner:

*Case I.* One boundary's first derivative is continuous at both points of intersection; the other boundary has a discontinuous first derivative at both points of intersection. This case is illustrated in figure 1 (*A*), in which the figure whose boundary has a continuous first derivative at the points of intersection, i.e., the left-hand figure, is seen as nearer; the figure whose boundary has a discontinuous first derivative, i.e., the right-hand figure, is seen as farther away. In figure 1 (*A*), the same cue is provided at both points of intersection.

*Case II.* One function possesses a continuous first derivative at one point of intersection and a discontinuous one at the other, while the other function's first derivative is discontinuous at both points of intersection. This case is shown in figure 1 (*B*). Here the point of intersection at which both first derivatives are discontinuous, point *a*, gives no cue. However, at point *b*, one function (that represented by the outline of the right-hand figure) has a discontinuous first derivative, and the other function (that represented by the outline of the left-hand figure) has a continuous first derivative. Thus, the left-hand figure is seen in front of the right-hand figure.

Cases I and II give familiar illustrations of interposition.

*Case III.* The first derivative of each function is continuous at one point of intersection and discontinuous at the other, and each derivative is continuous at that point of intersection at which the other is discontinuous. This case is illustrated in figure 1 (*C*). Here a cue is provided at each point of intersection, but the two cues contradict each other. The point of intersection at *c* indicates that the right-hand figure is nearer, while that at *d* indicates that the left-hand figure is nearer. Thus, the resulting cues are in conflict.

*Summary.*—Helmholtz's specification of interposition is reformulated. It is asserted that: (1) interposition can provide a cue only at the points where the outlines of two objects meet; and (2) the object whose contour possesses a continuous first derivative at those points will be seen as nearer than an adjacent object whose first derivative is discontinuous at those points. Some consequences of this formulation are discussed.

---

* This account was prepared under Project NR 142-404, Contract N6onr-271, Task Order IX, between Columbia University and the Office of Naval Research, U. S. Navy.

[1] Boring, E. G., *Sensation and Perception in the History of Experimental Psychology*, Appleton-Century, New York, 1942, p. 264.

[2] Dimmick, F. L., "Visual Space Perception," in Boring, E. G., Langfeld, H. S., and Weld, H. P. (Eds.), *Foundations of Psychology*, John Wiley, New York, 1948, p. 298.

[3] Munn, N. L., *Psychology*, Houghton Mifflin, New York, 1946, pp. 352–354.

[4] Brown, W., "The Perception of Spatial Relations," in Boring, E. G., Langfeld, H. S., and Weld, H. P. (Eds.), *Psychology*, John Wiley, New York, 1935, p. 216.

[5] Helmholtz, H. von, *Physiological Optics*, Vol. III, Optical Society of America, 1925, pp. 283–284.

[6] To be continuous at a point, the first derivative must be defined at that point. Accordingly, in the following examples given in figure 1, the coördinate axes must be taken in such a way that neither is parallel to the sides of the figures containing a point of intersection.

PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES
Volume 35 June 15, 1949 Number 6

---

# *VALUES AS A SUBJECT OF NATURAL SCIENCE INQUIRY*

By A. L. Kroeber

Department of Anthropology, Columbia University, New York

Retired, University of California

Read before the Academy, April 25, 1949

This essay maintains the proposition that the study of values is a proper and necessary part of the study of culture, viewed as an existing part of nature.

This is said not merely as proposal or program, but as a descriptive fact holding for much of actually existing practice in anthropology and the study of culture.

Whenever a cultural fact has significance, or historical reference, it also contains a value. Significance must be distinguished from cause—from that which made a cultural phenomenon happen or come to be. Significance must also be distinguished from the end or purpose served; and from organic needs, which in their turn can be resolved either into causes or into ends of culture phenomena.

That needs—also called drives, press, imperatives and such—exist, and that they underlie and precondition culture, is indubitable. It is also obvious that culture cannot be explained or derived from needs except very partially. Hunger has to be satisfied; but *how* it is satisfied by human beings can never be derived from their being hungry, nor from their specific bodily construction. Overwhelmingly the *how* can be understood only with reference to the remainder of the culture adhered to, present and past; modified somewhat—or preconditioned—by interaction with the opportunities afforded by natural environment. Moreover, large segments of culture only begin to operate, to come into being, after the primal needs have been satisfied, have had their tensions reduced or alleviated. Such are art, religion, science. Hence these segments cannot be explained at all from physiological needs.

The essential characteristic things about a culture are its forms and patterns, the interrelations of these into an organization and the way these parts, and the whole, work or function as a group of human beings lives under them. A culture is a way of habitual acting, feeling and thinking channeled by a society out of an infinite number and variety of potential ways of living. The particular channeling adopted is heavily preconditioned by antecedent ways and organizations or systems of culture; though it is not predetermined thereby except within certain limits. Every such system of channeling is accompanied by or contains a system of affects, which vary from place to place of their appearance, and from time to time, but some of which are usually powerful and persistent. Interconnected with these affects is a system of ideas and ideals, explicit and implicit. The combined affect-idea system of a culture at once reflects the habitual ways of action of members of the society, validates these ways to themselves and to an extent controls and modifies the ways. It is in this affect-laden idea system that, in a certain sense, the core of a culture is usually considered to reside: in it lodge its values, norms and standards—its ethos and its eidos.

When we speak of the significance of a cultural trait or item or complex of traits, what is meant is the degree to which the trait is meshed, affectively as well as structurally and functionally, to the remainder of the total system or organization that constitutes the culture. Low degree of integration normally indicates that the trait has relatively low significance for the culture as a functional unit—though it may still have considerable significance as an index of historical relationship with other cultures.

It follows that if we refuse to deal with values we are refusing to deal with what has most meaning in particular cultures as well as in human culture seen as a whole.

What we have left on elimination of values is an arid roster of cultural traits or cultural events which we are constantly tempted to animate by reintroducing the values we have banned, or else by backhandedly introducing values from our own culture. Or it is possible to attempt to explain the value-rid phenomena of the culture and their changes in terms of some causality; or possibly by a teleology.

As a matter of fact, it is and long has been prevailing practice in the description of cultures by anthropologists, or of civilizational phases by historians, to formulate the value of these cultures. Thereby the description becomes a physiognomic characterization of the culture. Such a characterization has internal import both as regards its own coherence and consistency, and external import through implicit or explicit comparison with other characterized cultures. This type of presentation, with clear-cut value designations, comprises all the most successful characterizations, or resynthesized analyses, of cultures, both by anthropologists such as Malinowski, Firth, Evans-Pritchard, and by non-professionals like Codrington on the Melanesians, Doughty in Arabia Deserta, Fustel de

Coulanges's Ancient City, Albiruni on India a thousand years ago and as far back as the Germania of Tacitus.

Reference in this matter is to values as they exist in human societies at given times and places; to values as they make their appearance in the history of our species; in short, to values as natural phenomena occurring in nature—much like the characteristic forms, qualities and abilities of animals as defined in comparative zoology. There is no reference to any absolute standard or scale of values; nor to judgments of values as better or worse—which would imply such a standard.

An absolute standard involves two qualities. First, it must be extra-natural, or supernatural, to be an *a priori* absolute. And second, ethnocentricity is implied in the elevation of any one actual standard as absolute. By contrary, standards or value-systems conceived as parts of nature are necessarily temporal and spatial, phenomenal, relative and comparative. That the first condition to the scientific study of culture is the barring of ethnocentrism has been a basic canon of anthropology for three-quarters of a century.

The forms of any culture must be described—can be appraised one might say—only in terms of their relation to the total pattern-system of that culture. The pattern system in its turn needs portrayal in terms of its total functioning and products. And so far as the pattern system is appraisable, it is in terms of comparison with the functions and results achieved by other total cultures with their respective master patterns. This is like the comparison of the total functioning and capacities of, say, an earthworm, with the functioning and capacities of other organisms.

In a sense, recognition of the functioning and capacities of an organic species in a sort of formulation of the values genetically inherent in that species. At any rate, it *can* be that, even if biologists usually are not aware of the fact and might resent the imputation of any concern with values. Further, the comparison of such values, in order to ascertain their common elements, their particularities, their apparent total range of variability, their effectivenesses and long-range permanences—such a comparison of biological values would still be within the scope of examination of natural phenomena by natural science methods.

It is as something analogous to this kind of biology or potential biology that the study of cultural forms, structures and values must be conceived. Or rather, we should say that such study has actually been made, time and again, often without explicit awareness of values being involved, and perhaps as often without awareness that the study had natural scientific significance.

It is true that values can also be viewed extra-scientifically or supernaturally. Mostly they have been so construed, with ascription to deity, soul, spirituality or a self-sufficient system of eternal, unmodifiable values lying outside the domain of science over nature. But the present paper has no concern with such a view. Contrarily, it claims values, along with all other manifestations of culture, as being part of nature and therefore in the field of science.

A few specifications seem desirable.

There is always a gap between values and behavior, between ideals and performance. Even though values always influence behavior of cultural organisms, that is, of men, they never control it exclusively. Hence the student of culture needs to distinguish, but also to compare, ideal values and achieved behavior, as complementary to each other. The one alone falls short in substantiality, the other in significant motivation and organization of the data.

Next, values being sociocultural, they inevitably also possess psychological aspects. But as a specific quality of culture, as indeed a product of culture, their reduction to explanations in psychological terms, and of these to physiological and biochemical explanations, necessarily loses or destroys the essential specific properties of the values. These are retainable in full only as long as the phenomena of value continue to be inspected on the cultural level.

Finally, since cultural phenomena are determined in several ways—inorganically by environment, organically, psychologically and socially, as well as by existent culture—it is evident that the causality of cultural phenomena is likely to be unusually complex. Moreover they lend themselves with very great difficulty to the isolation and simplification of experiment. Within culture itself, these considerations seem to apply with even more strength to values than to say artifact production or subsistence economy. Other things equal, a descriptive or historical approach would accordingly seem more readily fruitful, in scientific inquiry into values, than any searching for causes—even immediate causes. This statement is not to be construed as a methodological ban against the study of causality in values, but as an intimation that the causal approach is inherently difficult and that valid, non-speculative results bid fair to be thin and slow. While a formal approach is thus indicated as more fruitful, this need by no means be limited to enumerative description, nor to enumeration of sequences. Beyond these, the comparison of organization, functioning and interrelations of cultural values and value systems invite methodical scientific research.

# INDIVIDUAL METABOLIC PATTERNS, ALCOHOLISM, GENETOTROPHIC DISEASES*

BY ROGER J. WILLIAMS, L. JOE BERRY AND ERNEST BEERSTECHER, JR.

THE BIOCHEMICAL INSTITUTE AND THE DEPARTMENT OF CHEMISTRY OF THE UNIVERSITY OF TEXAS AND THE CLAYTON FOUNDATION FOR RESEARCH, AUSTIN

Read before the Academy, April 27, 1949

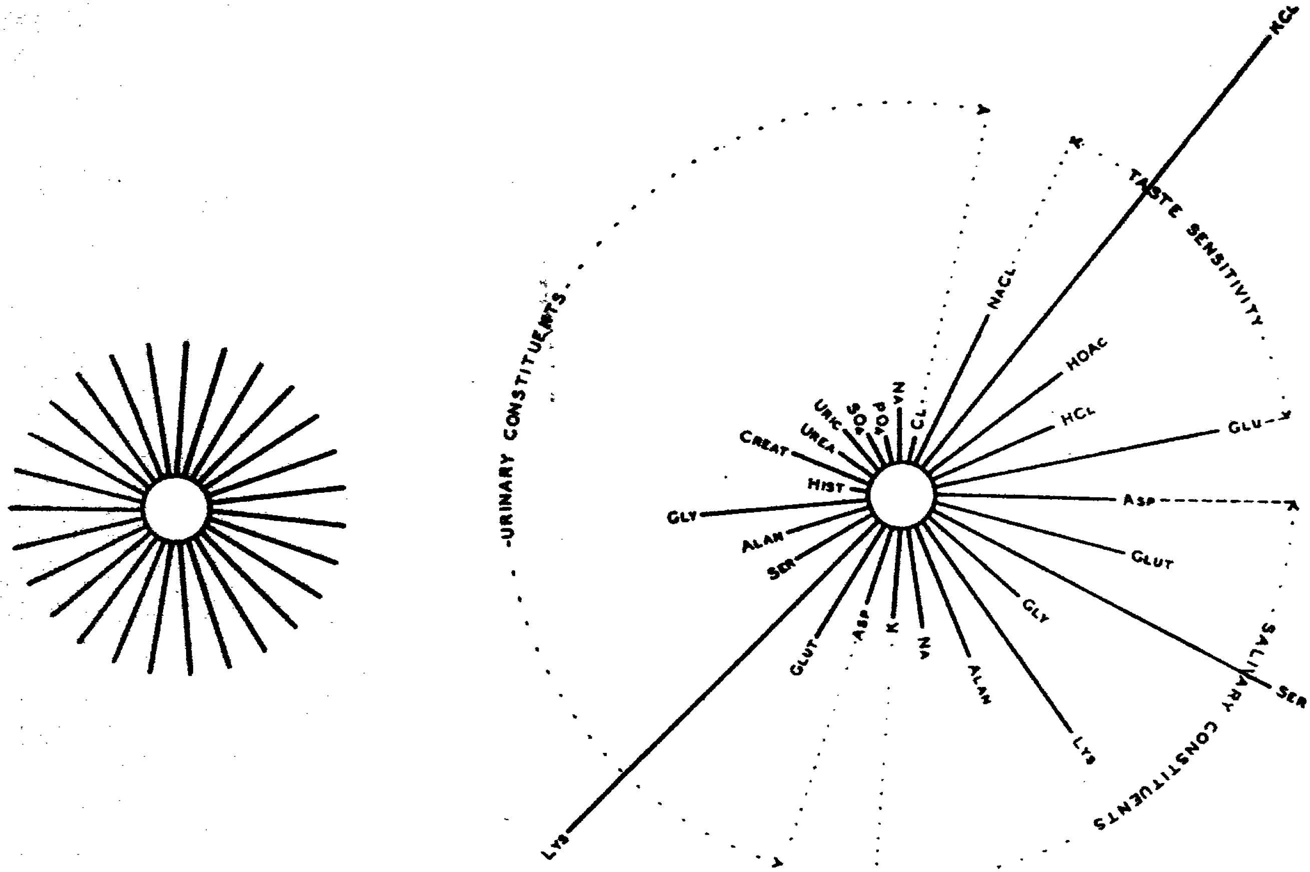
At the 1948 Annual Meeting of the National Academy there was presented preliminary evidence indicating that each individual possesses what may be called a "metabolic personality"—that is, a distinctive pattern of metabolic traits.[1] The existence of these distinctive patterns is established by analysis of body fluids and of physiological responses to chemical stimuli and is exemplified in figure 1. At the left is graphed, using polar coördinates, various metabolic traits of the "hypothetical average individual" whose every trait is the average of his fellows. At the right is graphed on the same scales the metabolic pattern of a typical real individual in which many traits vary widely from the average. This pattern is a relatively consistent one and is maintained for this individual at least over a period of several months.

The hypothesis was set forth that the metabolic patterns of individuals are of great import in the lives of the individuals who possess them, not only with respect to susceptibility to disease, but also with respect to nutrition, sex, mental abilities and many other facets of their lives.

Today we wish to present further and more specific evidence with respect to how these metabolic patterns are of moment in connection with the specific problem of alcoholism. This evidence is based primarily upon animal experimentation, the background of which is described elsewhere.[2-4]

It should be noted at the outset that individual animals in an ordinary laboratory colony have distinctive metabolic traits just as do the members of a human population. In order to conserve space we will not present data to substantiate this statement. When such animals are placed in individual cages and are continuously given a choice between water and 10% alcohol (the positions of the two drinking bottles being switched daily), they exhibit individual patterns with respect to alcohol consumption. Some drink fairly heavily beginning perhaps the first day the alcohol is offered. At the other extreme, some continue to abstain for an indefinite period of time. Intermediate between these are those that drink very moderately for an indefinite period of time and those that drink very little at first but after a few weeks drink relatively heavily. Some individual animals exhibit a relatively steady consumption day after day; in others the consumption fluctuates widely. (Figure 2)

These individual patterns of alcohol consumption are genetically con-



Hypothetical "Average" Individual

A Typical Individual

FIGURE 1

The following abbreviations are used in Figure 1: Creat = creatinine, Uric = uric acid, Hist = histidine, Gly = glycine, Alan = alanine, Ser = serine, Lys = lysine, Asp = aspartic acid, Glut = glutamic acid, Glu = glucose.

trolled as is evidenced by the fact that each strain of animals appears distinctive with respect to its alcohol consumption records. We have tested in a preliminary way three strains of mice and two strains of rats. Possibly more important evidence is the fact that when a closely inbred strain of mice resulting from the continuous use of brother-sister matings was tested, all the animals exhibited substantially the same pattern of response. Further convincing evidence on this point was obtained when the different strains of animals were tested while on different diets.

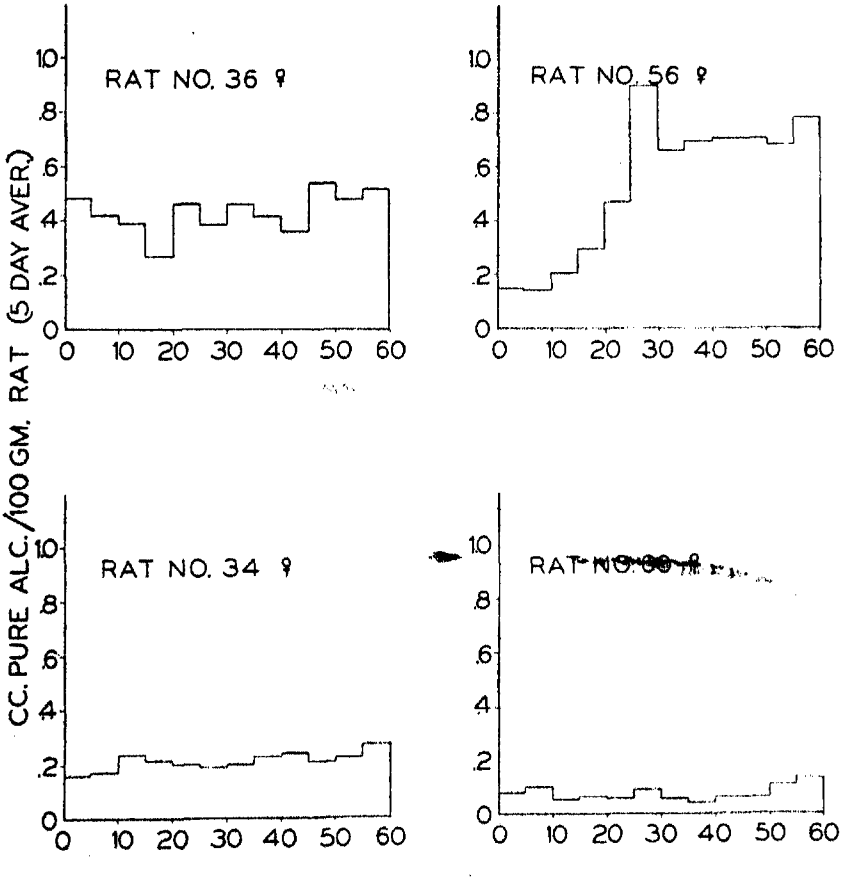


FIGURE 2

There has been a far-reaching and, we believe, disastrous tendency not only on the part of social students and educators in general but also to a degree among the medical profession to avoid the possible importance of heredity as a factor in any major human problem. The attitude can be expressed something like this: "If a certain disease is hereditary, or if mental traits and abilities are hereditary, nothing can be done about it. Therefore, let us assume that diseases are rarely hereditary and that every normal person is capable of unlimited development in almost any direction."

How wrong this attitude is will be amply illustrated if the results of our animal experiments are substantiated by human experience. Whereas our experiments have fully convinced us that the tendency of individual rats to consume alcohol is genetically controlled, this does not mean at all that nothing can be done about it. On the contrary, on the basis of somewhat unusual dietary considerations we have been able to abolish with surprising regularity the appetite of rats and mice for alcohol, even though it is clear that each animal has its genetically controlled metabolic peculiarities and that different factors in the treatment are effective for the different animals.

In Figures 3, 4 and 5 are depicted examples of the scores of rats which have been similarly affected. Rat No. 68 had his appetite for alcohol abolished by the simple oral administration of 10 known B vitamins. We do not know how many or which ones of the B vitamins were responsible for the effect in this particular rat. We do know, however, that only a small percentage of rats will yield to this treatment.

Rat No. 11, Figure 4, is one which responded, however, when in addition to the 10 B vitamins the antipernicious anemia vitamin was administered by injection. This treatment was found to be effective for 17 out of the 24 rats of our original laboratory strain, to which it was given. When the same treatment was given to 45 rats of a strain which we obtained from a neighboring laboratory, only 3, or 7%, responded. The consumption curve of rat No. 4, Figure 5, exemplifies the finding that for some rats vitamin A deficiency is a factor in the development of an appetite for alcohol. Thirty-one rats, whose consumption curves were relatively high, were administered vitamin A in addition to other supplements previously given, with the result that every single one showed immediately a decreased consumption.

The same curve shows that something in linseed oil—presumably linoleic and linolenic acids—is also a factor. Like a number of other rats, this one remained at a relatively high level of consumption even after the administration of 13 vitamins. When some linseed oil was included in the diet the consumption promptly fell to a low value.

Our experimental studies lead us to conclude that we can by dietary control keep all the rats in a colony from consuming alcohol when it is furnished *ad libitum* under our conditions. Even an initial experiment in this direction, made before we were acquainted with the effects of the antipernicious anemia vitamin, vitamin A or unsaturated fat acids, caused

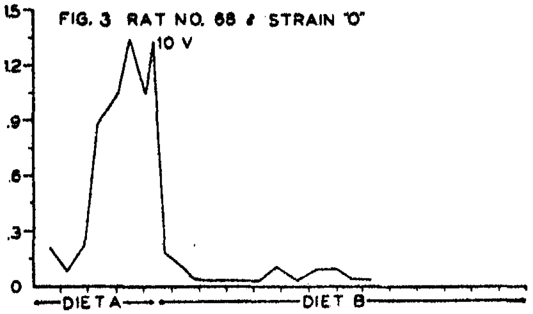


FIG. 3 RAT NO. 68 ♂ STRAIN "O"

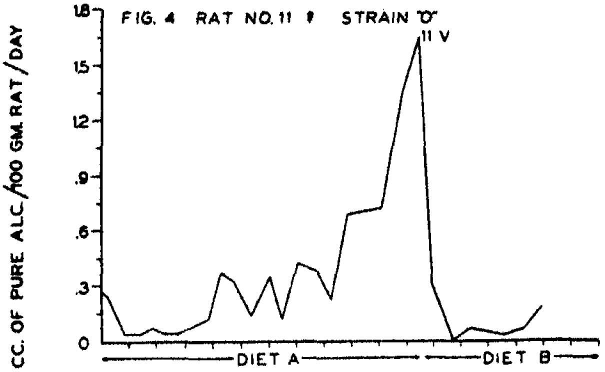


FIG. 4 RAT NO. 11 ♀ STRAIN "O"

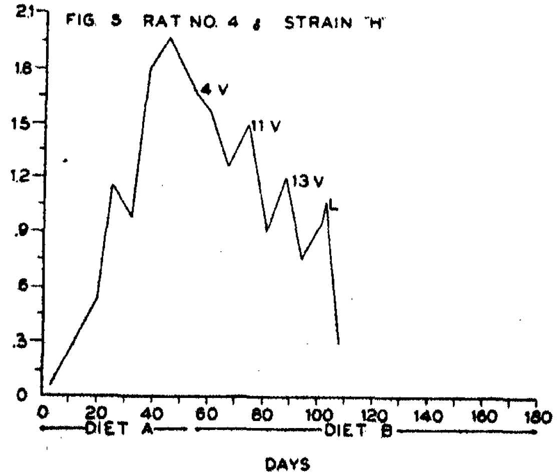


FIG. 5 RAT NO. 4 ♂ STRAIN "H"

Diet A is deficient in B vitamins. Diet B is the same except for added B vitamins. *4V* = thiamine, riboflavin, calcium pantothenate and folic acid (Folvite). *10V* = same as *4V* with addition of pyridoxine, choline, inositol, *p*-aminobenzoic acid, nicotinic acid and biotin. *11V* = *10V* with the addition of antipernicious anemia liver preparation given intraperitoneally. *13V* = *11V* with the addition of vitamins A and D. *L* = *13V* with the addition of linseed oil.

in a group of rats a decrease in consumption down to $^1/_7$ of the original.

We are also convinced that after an appetite for alcohol has been developed, it can be abolished in all rats by a use and extension of the method we have employed.

Out of these observations we have developed a theory of alcoholism which appears to be sound on the basis of all tests we have been able to apply. According to this theory, alcoholic craving which develops in compulsive drinkers constitutes a perverted appetite which arises as a result of one or more dietary deficiencies. Little is known as to how special appetites and cravings develop, but that they arise from the internal physiological environment has been amply demonstrated in a number of cases.

The dietary deficiencies in alcoholics arise not primarily because of failure to eat what is regarded as satisfactory food, but because the genetic patterns of these individuals are conducive to the development of crucial deficiencies. Partial genetic blocks which have recently been investigated at the California Institute of Technology[5] are thought to be responsible. When there is a partial genetic block, the capability of producing a specific enzyme is not entirely lost, but it is impaired to such an extent that there develops an augmented nutritional requirement which may be for a specific vitamin, amino acid, mineral element or other metabolic substance.

It is hoped that for the treatment of alcoholism in human beings a "shotgun" therapy may be developed which will effectively curb or abolish the excessive appetites of many compulsive drinkers.

Before treatment of alcoholics can be placed on a satisfactory basis, however, much more needs to be learned about individual metabolic patterns and what specific deficiencies are most often associated with the disease.

Out of this theory of alcoholism and the laboratory findings which support it we have developed the concept of genetotrophic disease (geneto = genetic; trophic = nutritional). A genetotrophic disease (of which we believe alcoholism is an example) is one arising fundamentally from nutritional deficiency which in turn has its basis in a genetically controlled augmented requirement for one or more specific nutritional elements. Other factors may enter to complicate or aggravate such a disease.

It suggests itself to us with great force that probably many diseases which have an obscure etiology may be essentially genetotrophic in origin. Among the diseases in which genetotrophic factors are probably involved in an important way are allergies, mental diseases, cardiovascular diseases, arthritis, multiple sclerosis and even cancer. This is hypothesis, of course, but it is based upon a large body of data.

Let us look for a few minutes at the tremendous problem of mental disease which has received wide-spread and well-deserved popular attention in recent years.

If we consider an example of a specific, important type of mental disease—schizophrenia—we are confronted by irresistable evidence that it has physiological concomitants. If we take a one-sided view, we will assume *a priori* that the physiological disturbances are psychogenic in origin and are a result of mental derangement. If we look at the problem more broadly, however, we will consider the possibility that mental derangement may be the result of physiological impairment. We believe there is abundant reason for investigating this latter possibility.

We have made a preliminary study of the metabolic patterns of a considerable number of patients of the Austin State Hospital. This study has revealed the existence among these patients of a number of metabolic traits which are either absent or rare in the well population. We are of the opinion that some of these traits will be found to be important in connection with the causation of mental diseases, and that such investigations need to be greatly extended.

The possibility that mental diseases are genetotrophic in origin or that they have genetotrophic factors in their etiology cannot be brushed aside lightly in view of the known fact that certain mental difficulties have been treated with remarkable success by physiological means. I refer to the severe mental difficulties that sometimes accompany menopause which can be completely dispelled by hormone therapy, and the mental derangement accompanying pellagra which can be cured by dietary means.

The recent survey by Kallmann[6] gives what appears to be practically irrefutable evidence that schizophrenia has a genetic basis. This by no means indicates that nothing can be done about it. When and if we find that certain metabolic difficulties are at the basis of the trouble, there is, we believe, an excellent chance that a way out can be found. If genetotrophic factors are found, as seems highly probable, certainly help can be given.

In any event the physiological approach to the study of mental disease needs far more emphasis than it has so far received. We cannot afford to turn the problem over to cultists or to those who are wholly unprepared for physiological exploration.

---

[*] This research was supported in part by grants from the Research Corporation and the Research Council on Problems of Alcohol, New York.

[1] Williams, R. J., *Science*, **107**, 459 (1948).

[2] Williams, R. J., Berry, L. J., and Beerstecher, E., Jr., *Arch. Biochem.* (in press).

[3] Mardones, J., Segovia, N., and Onfray, E., *Arch. Biochem.*, **9**, 401 (1946).

[4] Williams, R. J., *Chem. Eng. News*, **25**, 1112–1123 (1947).

[5] Mitchell, H. K., and Houlahan, M. B., *Fed. Proc.*, **6**, 506–509 (1947).

[6] Kallmann, F. J., "Personality, Society and Culture," (Kluckholn, C., and Murray, H. A., editors), A. H. Knopf, New York, N. Y., 1948, pp. 60–79.

# INHIBITION OF PLANT GROWTH BY PROTOANEMONIN AND COUMARIN, AND ITS PREVENTION BY BAL*

By Kenneth V. Thimann and Walter D. Bonner, Jr.

Biological Laboratories, Harvard University, Cambridge, Mass.

Read before the Academy, April 27, 1945

1. *Introduction*.—In two papers dealing with the growth by cell enlargement of sections of oat coleoptiles and pea stems, growing in simple solutions in the dark, the effects of several inhibitors were studied.[1, 2] It was deduced that the elongation of these isolated plant parts is mediated through a sulfhydryl enzyme, which is particularly open to attack by such substances as arsenite, iodoacetate and organic mercurials. These reagents are powerful growth inhibitors. In the case of iodoacetate the addition of malate or other organic acids can protect against the inhibitor, and there is other evidence that the sensitive enzyme is connected with organic acid metabolism. It also appears that this enzyme, or the reaction it catalyzes, is the main limiting factor in growth of these plants. Furthermore, it was tentatively suggested that the unsaturated lactones, which are well known to inhibit growth of seedlings, also do so through the same sulfhydryl enzyme. Although reasons were given for this view, no evidence was offered at the time. The present paper gives experimental evidence which, we believe, confirms the suggestion.

In the meantime a study of the literature has revealed evidence already in existence for an interaction between unsaturated lactones and thiol groups. Hauschka, Toennies and Swain[3] found that cysteine antagonized the action of δ-hexenolactone on the flat-worm *Dugesia tigrina*, and chemical evidence for a reversible reaction between the same lactone and cysteine has been given by Cavallito and Haskell.[4] Geiger and Conn[5] have shown that clavacin (patulin), an antibiotic of unsaturated lactone structure, is inactivated by cysteine and thioglycollate. This evidence strengthens the view proposed above, and points in the same direction as our own experiments on higher plants.

2. *The Inhibiting Action of Unsaturated Lactones on Growth*.—The inhibiting effect of a number of unsaturated lactones on seed germination and root elongation has been reported by many workers (see Audus and Quastel[6] and literature there cited). However, seed germination is an obscure process and appears to involve root elongation, which is auxin-sensitive, as well as the complex of enzyme activations participating in germination proper. Fortunately the lactones also exert a clear-cut inhibition on a well-studied growth process. This was first made clear by Veldstra and Havinga[7] using coumarin as the unsaturated lactone, and

studying the curvature of slit pea stems in auxin solutions. We have confirmed the effect on slit pea stems on several occasions (cf. reference 8); curvature is inhibited by concentrations of $1.10^{-4}$ *M* and above; the inhibition is complete, *i.e.*, the inward curvature is abolished, at about $3.10^{-3}$ *M*. The procedure of these experiments was as previously described.[9] As with arsenite (cf. reference 2, Fig. 5) the threshold concentration depends somewhat on the concentration of auxin used; *i.e.*, the higher the auxin concentration the larger the curvature of the uninhibited controls and the lower the concentration of coumarin which shows a detectable effect. The inhibition is not removed by adding malate or other organic acids.

Protoanemonin, the lactone of 3-hydroxy-$\Delta^{1,3}$-pentadienoic acid, which was kindly made available to us by Dr. Chester Stock, behaved much like coumarin but was effective at 10 to 30 times lower concentrations; inhibition was detectable at $4.10^{-6}$ *M* (0.4 p. p. m.) and complete at $1.10^{-4}$ *M*† (see table 1).

TABLE 1

INHIBITION OF SLIT PEA STEM CURVATURES BY PROTOANEMONIN. THE CURVATURES ARE MEASURED BY THE "STEM-REFERENCE" METHOD;[9] INWARD CURVATURE IS DESIGNATED AS POSITIVE, OUTWARD AS NEGATIVE

| CONCENTRATION OF INDOLE-ACETIC, MG./L. | CONCENTRATION OF PROTOANEMONIN, $\times 10^{-4}$ M | CURVATURES, DEGREES |
|---|---|---|
| 1 | 0 | +112 |
| 1 | 0.04 | + 87 |
| 1 | 0.2 | + 51 |
| 1 | 1.0 | −118 |
| 1 | 4.5 | Toxic |
| 0 | 0 | −116 |

Both lactones also inhibit growth of coleoptile sections, and do this at concentrations lower than those effective in the pea test. The technique of these experiments, using 3-mm. sections cut from 72-hour-old etiolated Avena coleoptiles, has been described in detail elsewhere.[1] Figure 1 summarizes the data of three experiments, one on coumarin and two on protoanemonin, using indole-acetic, 1 mg. per liter, and sucrose, 1%, as growth medium. To make the experiments comparable the growth of the control sections in the absence of lactone is set at 100%. The data show that low concentrations of both lactones accelerate growth, protoanemonin being more effective than coumarin. They also reveal that when inhibition begins to occur the remaining growth is linearly proportional to the logarithm of the concentration of lactone. The lines are so evidently straight that there can be no doubt of the relationship. The slope is different for the two lactones.

The shapes of these curves, with their marked growth promotion at low concentrations, are similar to those previously recorded for iodoacetate.[1] However, the inhibition is not prevented by malate or succinate. The effects of the lactones on growth are thus intermediate between those of iodoacetate and of arsenite.

It may be mentioned that Penicillin G, which has been thought also to react with SH groups, did not inhibit growth in concentrations up to $10^{-4}$ *M*.

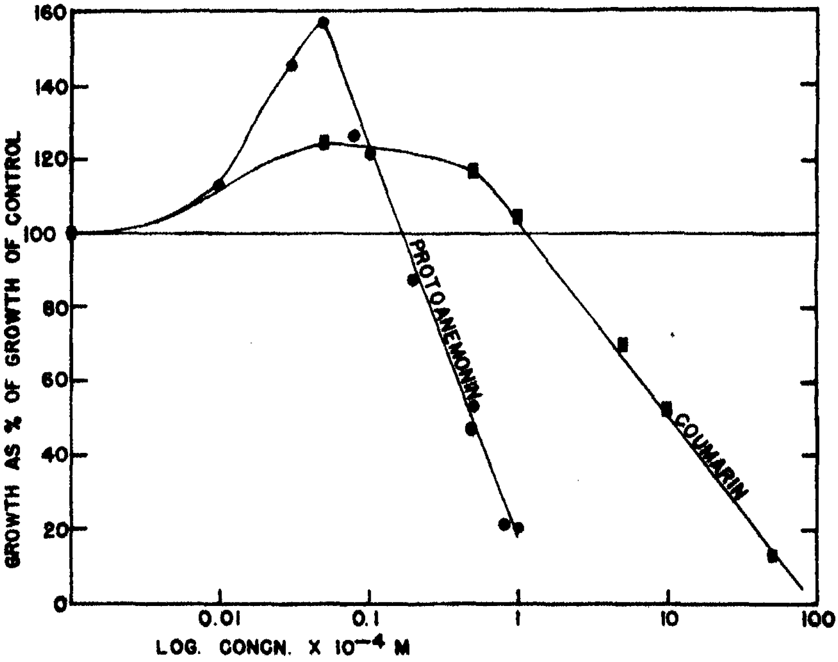


FIGURE 1

The effect of coumarin and protoanemonin on the growth of 3-mm. Avena coleoptile sections. The total growth after 48 hours, as per cent of the control growth, is plotted as a function of the log of the lactone concentration. Sections cut from 72-hour coleoptiles, breaking surface. All solutions contained sucrose, 1%, and 1 mg./l. indoleacetic acid.

*3. Protection Against the Inhibition.*—When the sections are treated with either one of the lactones and at the same time with 1,2-dimercaptopropane (BAL) the inhibition is very largely prevented. It has not been possible to achieve complete protection because the BAL itself has a toxic action above $3.10^{-4}$ *M*, but the effects are nevertheless quite clear-cut.

Table 2 shows an experiment with pea curvatures using protoanemonin at a concentration sufficient to give an inhibition of well over 50%. The BAL at $3.10^{-4}$ *M*, which by itself reduces the curvature 30°, is able to bring up the curvature in presence of protoanemonin from −39° to +121°.

Butanedithiol had a similar effect but the results were more erratic.

Table 3 shows a similar experiment using coumarin. The increases in curvature due to BAL are again very large, and in the first case bring the value above that of the controls. In this table the extent of the protection by a given concentration of BAL is more or less independent of the coumarin concentration.

With the growth of coleoptile sections, protection against coumarin was also real, but quantitatively less impressive. Coumarin at $10^{-3}$ *M* reduced the growth to 49% (cf. 53% in Fig. 1) and addition of BAL $1.10^{-4}$ *M* raised this to 61%. Table 4, however, shows excellent protection of coleoptile sections against the inhibition due to protoanemonin. It will be seen that growth which has been reduced to one-third that of the controls is increased to 88% of the control by BAL at $1.10^{-4}$ molal.

TABLE 2

INHIBITION OF CURVATURES OF SLIT PEA STEMS BY PROTOANEMONIN AND ITS PREVENTION BY BAL

| CONCENTRATION OF INDOLE-ACETIC, MG./L | CONCENTRATION OF PROTOANEMONIN, $\times 10^{-4}$ *M* | CONCENTRATION OF BAL, $\times 10^{-4}$ *M* | CURVATURE, DEGREES | INCREASE DUE TO BAL, DEGREES |
|---|---|---|---|---|
| 10 | 0 | 0 | +238 | ... |
| 10 | 0.4 | 0 | − 39 | ... |
| 10 | 0.4 | 0.5 | + 73 | +112 |
| 10 | 0.4 | 3 | +121 | +160 |
| 10 | 0 | 0.5 | +271 | ... |
| 10 | 0 | 3 | +207 | ... |
| 0 | 0 | 0 | −183 | ... |

TABLE 3

INHIBITION OF CURVATURE OF SLIT PEA STEMS BY COUMARIN, AND ITS PREVENTION BY BAL

| CONCENTRATION OF INDOLE-ACETIC, MG./L. | CONCENTRATION OF COUMARIN, $\times 10^{-4}$ *M* | CONCENTRATION OF BAL, $\times 10^{-4}$ *M* | CURVATURE, DEGREES | INCREASE DUE TO BAL, DEGREES |
|---|---|---|---|---|
| 10 | 0 | 0 | +234 | ... |
| 10 | 1 | 0 | +165 | ... |
| 10 | 1 | 3 | +290 | +125 |
| 10 | 3 | 0 | − 43 | ... |
| 10 | 3 | 3 | + 84 | +127 |
| 10 | 10 | 0 | − 65 | ... |
| 10 | 10 | 3 | − 34 | + 31 |
| 10 | 30 | 0 | −172 | ... |
| 10 | 30 | 3 | + 2 | +174 |
| 0 | 0 | 0 | −199 | ... |

When the inhibition is more nearly complete the protection against it by BAL or butanedithiol is quantitatively smaller. With peas in $10^{-4}$ *M* protoanemonin the effect of BAL was practically zero, and a similar reduction of effect is seen in the last two lines of table 4, in which, however, the protection is still definitely present.

4. *Summary and Conclusion.*—A study is reported of the action of coumarin and protoanemonin on the growth of two standard objects—sections of Avena coleoptiles elongating in auxin and sucrose, and slit internodes of Pisum stems curving in auxin alone. Both lactones inhibit these growth processes, the protoanemonin being 10 to 30 times the more active. At subinhibiting concentrations the growth is promoted in each case. The inhibition is clearly prevented by the addition of dimercaptopropanol, BAL, at concentrations below $3.10^{-4}$ molal.

TABLE 4

INHIBITION OF GROWTH OF 3-MM. COLEOPTILE SECTIONS BY PROTOANEMONIN AND ITS PREVENTION BY BAL. ALL SOLUTIONS CONTAIN INDOLE-ACETIC ACID 1 MG./L. AND SUCROSE 1%

| CONCENTRATION OF PROTOANEMONIN, $\times 10^{-5}$ $M$ | CONCENTRATION OF BAL, $\times 10^{-4}$ $M$ | GROWTH AS PER CENT INCREASE OVER THE INITIAL LENGTH | GROWTH AS PER CENT OF GROWTH OF CONTROLS | INCREASE DUE TO BAL, % |
|---|---|---|---|---|
| 0 | 0 | 91.0 | (100) | .. |
| 0.5 | 0 | 30.1 | 33 | .. |
| 0.5 | 1 | 79.9 | 88 | +55 |
| 0.5 | 3 | 71.3 | 78 | +45 |
| 1.0 | 0 | 14.6 | 16 | ... |
| 1.0 | 3 | 31.8 | 35 | +19 |

It is concluded that coumarin and protoanemonin inhibit growth through reacting with a sulfhydryl enzyme, and that this enzyme is normally a limiting factor in growth, and is very probably the same as that inhibited by iodoacetate and arsenite.

[*] This work was made possible through a grant from the Committee on Growth of the National Research Council, acting for the American Cancer Society.

[†] The instability of the lactone in aqueous solution, even at 1 °C., limited the number of possible experiments.

[1] Thimann, K. V., and Bonner, W. D., I, *Am. J. Bot.*, **35**, No. 5, 271–281 (1948).

[2] Thimann, K. V., and Bonner, W. D., II, *Ibid.*, **36**, No. 2, 214 (1949).

[3] Hauschka, T., Toennies, G., and Swain, A. P., *Science*, **101**, 383 (1945).

[4] Cavallito, C. J., and Haskell, T. H., *J. Am. Chem. Soc.*, **67**, 1991 (1945).

[5] Geiger, W. B., and Conn, J. E., *Ibid.*, **67**, 112 (1945).

[6] Audus, L. J., and Quastel, J. H., *Nature*, **159**, 320 (1947).

[7] Veldstra, H., and Havinga, E., *Rec. trav. chim.*, **62**, 841 (1943).

[8] Leopold, A. C., and Thimann, K. V., *Am. J. Bot.*, **36**, No. 4, 342 (1949).

[9] Thimann, K. V., and Schneider, C. L., *Ibid.*, **25**, No. 8, 627 (1938).

# DEVELOPMENTAL ANALYSIS OF GENETIC DIFFERENCES IN PIGMENTATION IN THE AXOLOTL

By H. Clark Dalton

Department of Genetics, Carnegie Institution of Washington, Cold Spring Harbor, N. Y.

Communicated by M. Demerec, April 4, 1949

The pattern differences in pigmentation exhibited by the white and black strains of the Mexican axolotl, *Siredon mexicanum*, were demonstrated by Haecker[1] in 1907 to be due to a single pair of alleles showing Mendelian segregation of the white condition as the recessive homozygote. Investigations on the nature of gene action in this case have been limited. Woronzowa,[2, 3] implanting pituitaries in postlarval stages, observed increases in pigmentation, and concluded from the differences in dosage required by the two strains that the recessive gene when homozygous raises the threshold of the skin to pituitary stimulation. Some information on the early developmental action of this gene has come from embryonic transplantations. The results of these experiments, summarized by DuShane,[4] have been interpreted as indicating that the genetic pigmentation differences between the two strains are mediated through the epidermis, which in the white genotype fails to provide some diffusible substance (oxidase or chromogen) necessary for melanin synthesis. The experiments here reported investigate further the developmental basis for this genetic difference in pigment pattern, by methods of tissue culture and embryonic transplantation.

*Explantation of Neural Fold.*—The pigment-forming potentialities of melanophores of the two genotypes, *DD* (black strain) and *dd* (white strain), were tested by growing explants of neural fold from the posterior trunk region in Holtfreter's solution in hanging-drop preparations (method of Twitty[5]). The cultures were maintained in a constant-temperature room at 18°C. The results have bearing on two questions: (1) Is the effect of gene *d* on the melanophores themselves or on their surroundings in the embryo? (2) Are the melanoblasts dependent on the epidermis for a substance necessary for melanin synthesis? Cultures of both genotypes showed the same characteristics of development. Melanoblasts began migrating out from the explants onto the cover glass in two or three days. In five days many cells had assumed the branched form characteristic of chromatophores, and some cells appeared full of gray granules. During the next four or five days more and more cells appeared full of granules, which darkened to become the black melanin granules found in differentiated melanophores. The development of melanophores in Holtfreter's solution is limited by the amount of nutrient reserves within the cells. With the disappearance in about two weeks of the yolk platelets, which are large and numerous in early melanoblasts, development seemed to stop, although some cultures remained apparently healthy for as long as thirty-eight days. The developing time was long enough, however, to allow pigment production to the extent that many of the cells became very dark gray, and some black. Xanthophores were discernible, in cultures of both genotypes, one or two days after the melanophores appeared full of granules. No difference was found between cells of white and black strains with respect to extent of outgrowth or degree of pigmentation attained by melanophores. This result gives direct support to the idea, previously drawn from transplantation experiments, that the gene *d* acts on the environment of the chromatophores rather than on the chromatophores themselves. Since melanophores of both genotypes produced pigment in their inorganic medium, there is in this experiment no evidence that the melanoblasts of this species are dependent on a substance from the epidermis for melanin synthesis.

*Embyronic Transplantations.*—Information about the relationships between developing chromatophores and their surrounding tissues has been obtained by arranging these components of the two genotypes in different combinations. An experiment was planned to answer the question, do pigment cells of the same origin differ in development when in association with tissues of the black strain and of the white strain? Neural folds from the posterior trunk region of donors corresponding to Harrison stages 18–20 were transplanted to the midventral belly region of hosts in stages 22–23. Each donor provided one transplant in a black host and one in a white host, allowing subsequent comparison of graft chromatophores of identical genotype and age developing in the two environments. This transplantation site has the advantage of obviating any confusion of graft chromatophores with those of the host, since the ventral belly region does not become pigmented in either strain during the period of observation of this experiment. Six pairs of such transplants were made, three from black donors and three from white donors, with one extra case of a black-to-white graft. In all cases melanoblasts migrated out from the graft and became associated with host tissues. Two paths of migration were available to them; since the operation resulted in all cases in adhesions between skin and gut at the graft site, migrating cells could move out from the graft on either the skin side or the gut side. Pigment appeared in graft melanophores after 4 days in twelve cases and after 3 days in one case of black-to-white. Thus no delay was evident in the time of pigmentation of cells growing in the white hosts. With regard to degree of pigmentation, all melanophores became completely black in both types of host. The only apparent difference observed was that the melanophores of black origin in all cases appeared darker for a few days in the white hosts than in the black hosts. This distinction was temporary, and does not necessarily reflect differences in the amount of pigment present, since variations in shape and orientation of the cells might easily give this appearance. From these direct comparisons it must be concluded that melanophores growing in association with tissues of the white strain reflect no disadvantage for melanin synthesis, in terms of differences in rate or degree of pigment formation, when compared with cells of the same origin growing in hosts of the black strain.

Comparison of the extent of migration of melanophores in the two types of host, on the other hand, revealed a striking difference. It was apparent from examination of the living animals that from all six grafts in black hosts melanophores had moved out extensively in both skin and gut tissues, whereas in four of the six grafts on white hosts the melanophores were restricted to the area of the graft itself and to the gut, none migrating into the skin. To obtain a more precise measure of the relative migration in the two types of host the following procedure was carried out. The animals were fixed in Zenker's fluid, four at 13 days and eight at 26 days after the operation. The skin was dissected off from the ventral side, cleared, and mounted in balsam; and the gut, from the stomach to the cloaca, was dissected out for observation in alcohol in a Syracuse dish. Camera lucida drawings were made of the areas occupied by donor melanophores in the skin and on the gut. Two drawings were made of each gut, to include both

TABLE 1

AREA IN SQUARE MILLIMETERS OCCUPIED BY MELANOPHORES FROM PAIRED BELLY GRAFTS

| | BLACK HOST | WHITE HOST |
|---|---|---|
| Black donor | 4.23 | 0.91 |
| | 1.35 | 0.59 |
| | 1.51 | 0.40 |
| | .. | 0.37 |
| White donor | 3.70 | 0.77 |
| | 8.73 | .. |
| | 4.01 | 0.52 |

TABLE 2

TOTAL NUMBER OF MELANOPHORES FROM PAIRED BELLY GRAFTS

| | BLACK HOST | WHITE HOST |
|---|---|---|
| Black donor | 452 | 119 |
| | 143 | 124 |
| | 57 | 76 |
| | ... | 52 |
| White donor | 258 | 108 |
| | 341 | ... |
| | 248 | 89 |

sides. Donor melanophores migrating on the gut of black hosts surrounded the gut, in all six cases, and covered extensive areas anterior and posterior to the graft, whereas donor melanophores migrating on the gut of white hosts were in all six cases restricted to the side of the gut contacting the graft, and covered less area. The areas occupied by melanophores in the drawings were measured with a planimeter, and the figures converted to square millimeters. No account was taken of the curvature of the gut surface, the whole procedure being at best an approximation. It is apparent from Table 1 that the total area occupied by graft melanophores, regardless of donor type, is much greater in black hosts than in white. The average area covered in black hosts was 3.92 mm.$^2$, compared with 0.59 mm.$^2$ in white hosts. The smaller areas observed in white hosts might have been due to reduced numbers of melanophores as well as to impairment of their migratory activities. To check this possibility, melanophore counts were made in the skin and gut preparations. Results are presented in table 2, where it is evident that in all but one case more melanophores developed from grafts in black hosts than from corresponding grafts in white hosts. The difference in areas cannot be accounted for, however, by the relative numbers of cells from the grafts. When one calculates the number of pigment cells per mm.$^2$ in each case, it is obvious that the melanophores in the white hosts are considerably more crowded than those in the black hosts. The average count for donor melanophores on black hosts was 70 cells per mm.$^2$, compared with 164 cells per mm.$^2$ on the white hosts. The results are shown graphically in figure 1. These measurements lead to the conclusion that the melanoblasts in white hosts are subject to some inhibition of movement.

The migratory inhibition observed in the foregoing experiment was manifest in a region of the body which in the white strain never becomes pigmented and in the black strain is pigmented only after considerable delay in comparison with other areas. It may well be asked, therefore, whether there are special conditions in the ventral belly region which influence pigmentation there but have nothing to do with development of the normal pigment pattern dorsally. To test the development of melanophores of each genotype growing in a tissue environment of the other at the normal position in the body, reciprocal neural crest grafts were made between pairs of the two types of embryos in stages 23–27. Transplantation of the neural crest necessarily involves transfer also of the dorsal epidermis overlying it. In order to obtain graft melanophores growing in complete absence of any donor epidermis—a factor of some importance if there is a question of diffusible substances—the operation was made in two steps. First, for example, a piece of dorsal epidermis overlying the neural crest of a black embryo was replaced by a piece of flank epidermis from a white embryo. Flank, rather than dorsal, epidermis was chosen to prevent the possibility of transplanting any adhering neural crest cells. On the following day the black neural crest, covered now by white epidermis, was transplanted to a white embryo. Thus, donor melanophores could differentiate in association with epidermis of host type only. By this method five transplants of black neural crest to white hosts and four transplants of white neural crest to black hosts were made. For comparison, a few neural crest transplants bearing donor-type epidermis were also made (two black to white, three white to black). The experiment is not yet terminated, since eight of the animals are still growing, but the effect on larval pigmentation was clear and consistent for each type of transplant.

When black neural crest covered by white flank epidermis was grafted to a white host, a dense black line of melanophores appeared beneath the epidermis at the edge of the graft region, which was clearly indicated by a discrepancy in the dorsal fin in that area. When black epidermis was grafted along with the neural crest to a white host, the dorsal fin developed intact and graft melanophores migrated abundantly into the donor portion of the fin but not into the adjacent host portion of the fin. The ventral limit of donor melanophores was not a sharp border, as in the cases described above, but the few large melanophores appearing below the graft remained on the dorsal part of the flank. These results suggest that the mobility of *DD* melanoblasts is greater in an environment of *DD* tissue than in one of *dd* tissue.

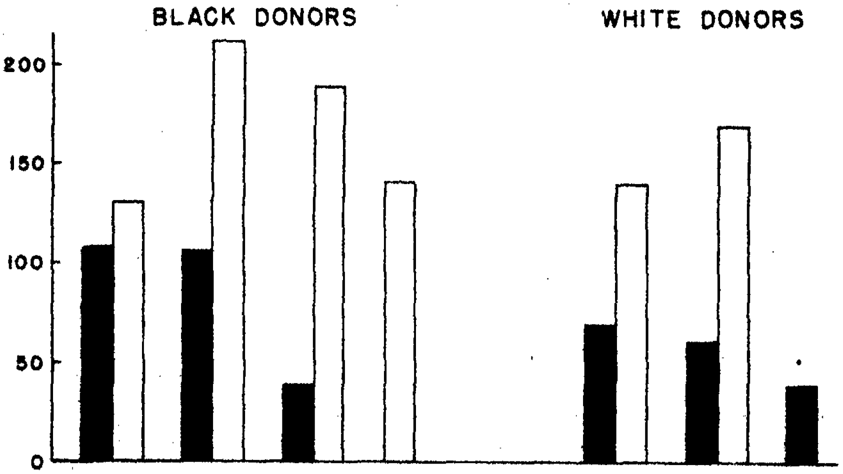


FIGURE 1

Number of melanophores per square millimeter from paired belly grafts in black hosts (solid) and in white hosts (unshaded). Adjacent columns represent grafts from the same donor.

When white neural crest covered by black flank epidermis was grafted to a black host, the graft region was later indicated, as in the reciprocal experiment, by a discrepancy in the dorsal fin. Donor melanophores appeared widely dispersed on the flank under the graft, and in two cases in extensive areas posterior to the graft region as well. These melanophores appeared somewhat smaller than the host melanophores, and did not form by overlapping the large irregular spots characteristic of the host pigment pattern at this stage. A further indication that the area of the flank beneath the grafts was inhabited by donor chromatophores came from the xanthophores, which in the white animals used in these experiments were much deeper yellow than those of the black animals. This distinction was very striking in black hosts bearing white neural crest grafts, the deep yellow areas occupied by donor xanthophores being visible to the naked eye. The donor melanophores and xanthophores migrating from grafts of white neural crest covered by donor epidermis on black animals showed the same wide dispersal on the flank of the host as was noted in the transplantations just described. These results suggest that the mobility of *dd* melanoblasts also is greater in *DD* tissues than it is in *dd* tissue. The evidence from all these homotopic neural crest transplantations indicates that the phenomenon observed in the belly grafts is not a special case, but that conditions affecting melanophore migration in the normal environment of the developing pigment cells are different in the two genotypes and are significant in the formation of the genetic pigment patterns characteristic of the two strains.

*Discussion.*—On the basis of the hypothesis that the gene *d* affects the epidermis of the recessive homozygote in such a way that it does not supply to the underlying melanoblasts a substance necessary for pigment production, which is, however, furnished by epidermis containing the allele *D*, one would have expected that *in vitro* in Holtfreter's solution melanoblasts of both genotypes would fail to produce melanin. That they both did become pigmented argues against this hypothesis. Some epidermal cells were undoubtedly explanted with the neural folds, but if these were responsible for allowing pigment formation in the melanophores, there should have been a difference in this respect between explants of the two strains, whereas no such difference was observed. To account for the presence of pigment cells in normal white animals, DuShane[6] suggested that there may be some melanophores that are independent of the epidermis effect. If the melanophores that developed from explants of white neural fold represented such independent pigment cells, one would have expected to find in addition many melanophores that remained colorless. A few non-pigmented cells otherwise morphologically resembling chromatophores were observed, but they were few in number (less than 4%) and occurred with the same frequency in explants from both strains. The idea of diffusible substances from the epidermis as the mechanism of gene action in this case is also inapplicable to the results of the transplantation experiments, where melanin synthesis was carried out equally well by melanoblasts in both types of tissue environment. This hypothesis concerning the nature of the epidermis effect is based on the assumption that melanoblasts migrate in the white axolotl just as they do in the black strain and in other pigmented species that have been used for reciprocal transplantations. This assumption does not explain the larval pattern distribution of melanophores in the white strain; furthermore, the impairment of mobility of melanoblasts in *dd* tissues reported here suggests that it may not be valid. The results of the present transplantation experiments, as well as the larval pigment pattern differences between the white and black strains of axolotl, are easily explained by the hypothesis that the melanoblasts of both genotypes are capable of producing pigment but that a reaction between melanoblasts and surrounding tissues of the white strain inhibits melanoblast migration. Whether this reaction is physical or chemical in nature cannot be stated at this time. Further experiments to obtain critical evidence on the problem are in progress.

*Summary.*—The differentiation of pigment cells from explanted neural folds of black and white axolotls in Holtfreter's solution has shown that potential melanophores of both genotypes are capable of melanin synthesis *in vitro.* Melanophores from grafts of neural folds to the midventral belly region of both dark and white hosts exhibit no disadvantages for melanin synthesis in white embryos. Chromatophores from reciprocal grafts of neural crest between white and dark strains also produce pigment in both types of host. A striking difference in the distribution of melanophores according to host genotype has appeared in both homotopic and heterotopic grafts. Melanoblasts of both genotypes appear to migrate freely beneath epidermis of the black strain but not beneath epidermis of the white strain. The results of all experiments support the idea that the genetic differences in pigmentation are mediated through differences in the tissue environment and not in the chromatophores themselves, the white pattern resulting from an inhibitory effect concerned not with conditions for melanin synthesis but with the migration of melanoblasts.

I wish to express my appreciation to Dr. E. W. Caspari for his continuous advice and encouragement during the course of this investigation.

---

[1] Haecker, V., *Zool. Anz.*, **31**, 99–102 (1907).

[2] Woronzowa, M. A., *Trans. Lab. Exp. Biol. Zoopark Moscow*, **4**, 107–124 (1928).

[3] Woronzowa, M. A., *Arch. Entw. Mech. Org.*, **115**, 93–109 (1929).

[4] DuShane, G. P., *Spec. Pub. N. Y. Acad. Sci.*, **4**, 1–14 (1948).

[5] Twitty, V. C., *J. Exp. Zool.*, **100**, 141–178 (1945).

[6] DuShane, G. P., *Ibid.*, **72**, 1–31 (1935).

# INHERITANCE OF CAROTENOID DIFFERENCES IN LYCOPERSICON ESCULENTUM STRAINS

By G. MacKinney and J. A. Jenkins

Divisions of Food Technology and Genetics, University of California, Berkeley, California

Communicated by E. B. Babcock, April 1, 1949

Two recessive mutant types of the common red-fruited tomato have long been known, the tangerine and the yellow. MacArthur[1] established the fact that these two color variants are due to recessive alleles at two loci situated on different chromosomes. They are therefore inherited independently. Using MacArthur's symbols, the genotypic formulae of the three different color types with respect to the two loci are: red *RRTT*, yellow *rrTT* and tangerine *RRtt*.

Ultimately, we hope to evaluate quantitatively the various combinations, such as, for example, those producing red fruit, namely *RRTT*, *RRTt*, *RrTT* and *RrTt*. The precise rôle of *R* and *T* genes cannot be established until the double recessive *rrtt* has been described. So far as we are aware, this has not yet been described, though we hope to identify it shortly.

In this paper we present evidence identifying the major carotenoid components for the three homozygous lines, and the relative proportions in which they are present. Data assembled over the past three seasons confirm and extend much that has been published[2–4] though some discrepancies exist. Pigmentation in a given phenotype is the expression of two factors, genetic constitution and environment. Lycopene formation is particularly sensitive to temperature. Thus, the fact that some twenty pure lines of yellow tomatoes studied here (differing greatly in pigment content and in other respects) invariably contained small quantities of lycopene, does not invalidate the argument[2] that the *R* gene is required (*inter alia*) for its normal production. In the absence of *R* the major and normal line of synthesis is blocked, but this need not preclude the feasibility of lycopene formation, under certain environmental conditions, in small amount, possibly with greater difficulty, by another path.

The red tomato owes its characteristic color to lycopene, the orange-colored tangerine to prolycopene, as noted by Zechmeister *et al.*,[3] and the yellow tomato to $\beta$-carotene and similar components. The three types are so characteristic that an examination of the absorption spectrum of a crude extract suffices to determine the type. Characteristic absorption curves are shown in figure 1. They are remarkably constant in shape for a given strain.

It is particularly interesting that while prolycopene determines the color of the tangerine tomato, its absorption is dominated (curves IV and V, figure 1) by pigments with maxima between 430 and 400 m$\mu$. This is

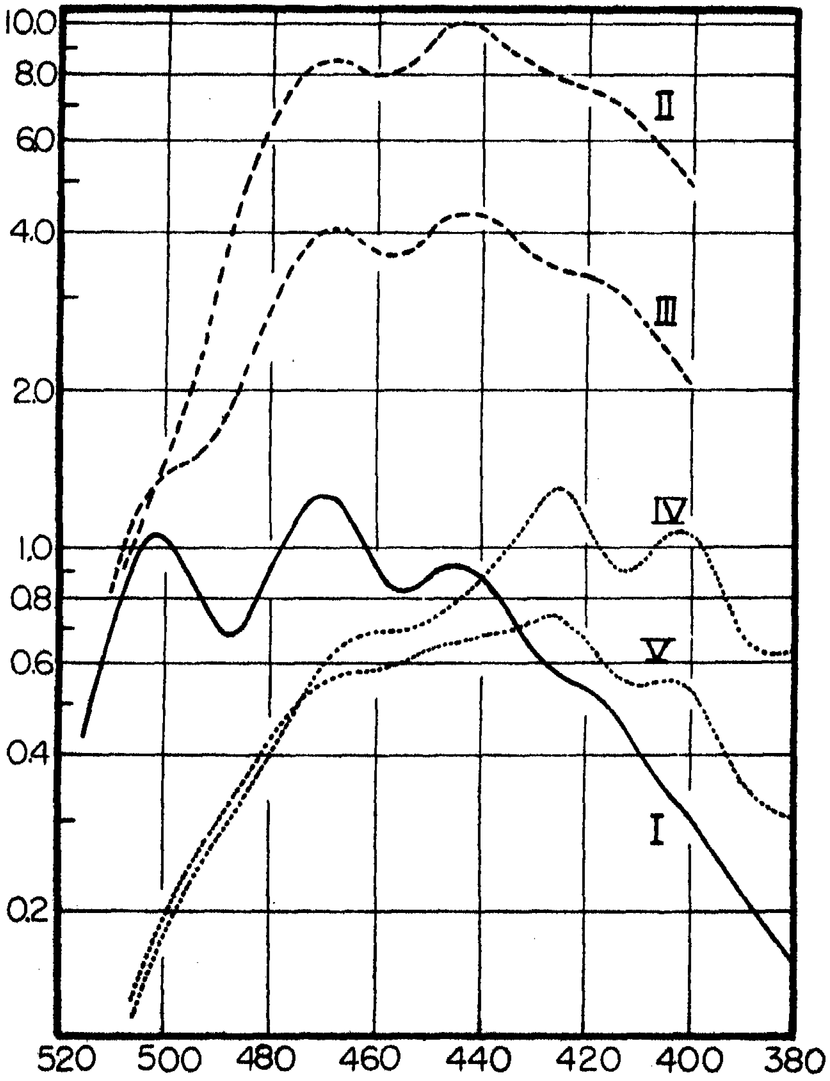


FIGURE 1

Absorption curves for typical extracts. Ordinate: optical density in arbitrary units; abscissa: wave-length. Curve 1 red, *RRTT;* curves II and III yellow, *rrTT,* curve III has more lycopene than curve II; curves IV and V tangerine *RRtt.*

accounted for by two pigments, all-trans ζ-carotene, and the "unidentified carotene I" of Porter and Zscheile.[4] For convenience in reference, it is tentatively designated as a poly-cis ψ-carotene. It has an all-trans spectrum apparently identical with Porter and Zscheile's "unidentified carotene II,"[4] but there is some reason for doubting whether their carotene II is a member of the ψ-set. Traces of other cis ψ-carotenes are present in the tangerine, and possibly also in the yellow tomato. They are readily separable, but are present in the yellow tomato in minute quantity, sufficient only for comparison of iodine-catalyzed spectra, indistinguishable from that for poly-cis ψ-carotene. The absorption curves for poly-cis ψ-carotene and its iodine equilibrium mixture are shown in figure 2. As surmized by Porter and Zscheile,[4] it cannot be a poly-cis lycopene. In so far as pure lines homozygous with respect to the *R* and *T* loci are concerned, poly-cis ψ-carotene appears to be unique to the tangerine. It is found in quantity (cf. [4, 5]) on the Tswett column as a bright yellow zone below prolycopene, above ζ-carotene. It is therefore spectroscopically and chromatographically comparable with "carotene I" of Porter and Zscheile. But for different spectroscopic results reported by Zechmeister and Le Rosen,[5] we should also consider it to be identical with zone 16 of their table 2. Attempts to crystallize poly-cis ψ-carotene have not yet been successful. Repeated chromatograms have failed to eliminate the shelf, *ca* 455 mμ in the poly-cis spectrum, or to change its height relative to the height at 430 mμ.

*Experimental.—Materials:** Twenty-two pure lines of yellow tomato, six of tangerine and three red have been examined, several repeatedly, to assess variability.

*Preparation of Extracts:* To minimize possible change, fruits were extracted rapidly, and measurements made immediately. Normally about 100 g. fresh weight of tomato from a single plant was blended in 300 ml. acetone, filtered on a Buchner and the residue re-suspended and the blending repeated.

In the case of the yellow and tangerine tomatoes the fleshy residue was white after a second filtration, but the skins were yellow if non-carotenoid alkali soluble skin-pigment was present. Where the red tomato contained carotenoid skin pigment, grinding with sand and a small quantity of petroleum ether was also necessary. . Three such treatments gave a fairly colorless residue.

Successive portions of the aqueous acetone extract were then transferred to 100 ml. petroleum ether, and water added to remove acetone, until all pigment had been transferred to the petroleum ether.

The petroleum ether solution was then freed from acetone by washing, and dried with anhydrous sodium sulfate.

*Absorption Spectra:* The absorption spectra of suitably diluted aliquots in petroleum ether were immediately determined on the Beckman DU spectrophotometer, tungsten lamp, in the range 520 to 330 mμ at 5-mμ intervals.

The curves shown in figure 1 are typical of more than 100 analyses. Apart from the absolute quantity of pigment produced, a given yellow strain shows greatest variability with respect to its lycopene content. This is reflected in the absorption at 500 mμ, relative to other wave-lengths, illustrated in curves II and III of figure 1. They were obtained from extracts of tomatoes taken from the same vine, at the same time. Those represented in curve III had pale pinkish streaks. Lycopene was also detected in the extract represented in curve II, but in lesser amount.

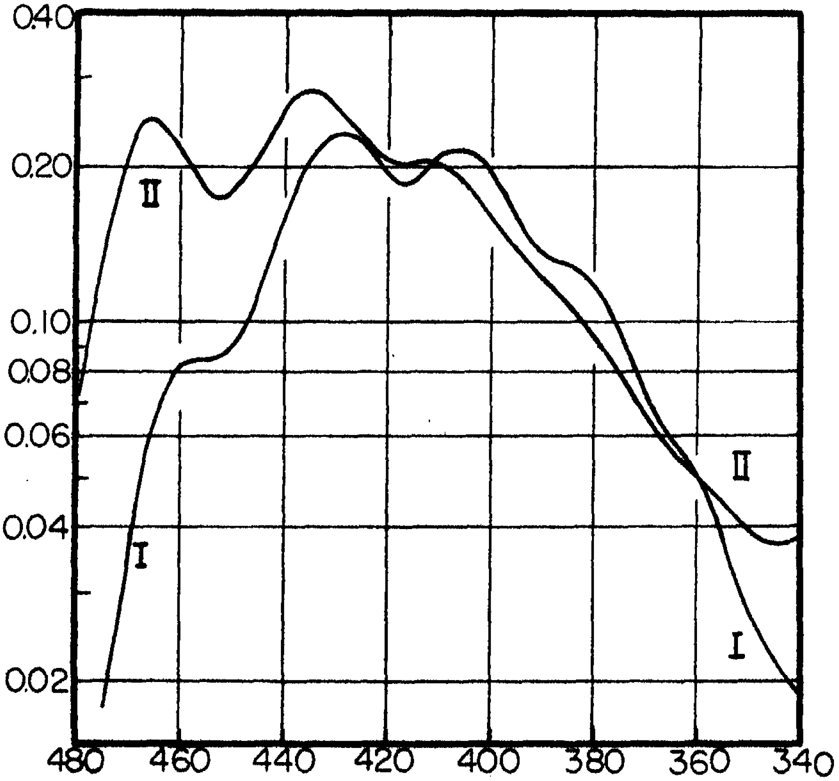


FIGURE 2

Absorption curves for ψ-carotene before (curve I) and after (curve II) iodine catalysis. The curves are directly comparable with respect to concentration, and therefore to optical densities.

Red tomatoes varied in carotene content but the spectrum was invariably dominated by lycopene (curve I). Tangerine extracts fell into two categories, consistently reproducible over the period during which fruit could be harvested. Curves IV and V represent extremes in the strains of this line.

*The Individual Components:* Petroleum ether extracts, freshly prepared as above, were chromatographed on magnesium oxide-hyflo-super cel (1:2). Columns packed to a height of 25 cm. and 2 cm. diameter were used. The more weakly adsorbed members were separated with this solvent, and identified. For strongly adsorbed components, fresh columns of adsorbent were used, the solvent being petroleum ether containing from 5 up to as much as 15% acetone. Before subsequent spectroscopic examination, the acetone was removed, in the usual way. As recommended by Zechmeister, an iodine catalysis was routinely included in the run, the two sets of spectrophotometric data being made simultaneously on aliquots from the parent solution.

In view of available detailed reports[2–5], brief comment will suffice except where discrepancies exist.

*Yellow Tomatoes:* Phytofluene, barely detectable in trace amounts, small bands at 370 and 348 being just discerned; the β-carotene set showed maxima unchanged by iodine catalysis; ζ-carotene could not be detected; minor components with well-separated bands follow next, giving a common iodine-catalyzed spectrum indistinguishable from that for poly-cis ψ-carotene; all-trans lycopene was noted, max. 505, 470, 445; shifting to 500, 468, 445 mμ on addition of iodine.

*Tangerine Tomatoes:* The detail reported by Zechmeister and Le Rosen[5] (table 2) is confirmed except for zones 15 and 16, neither of which give the lycopene spectrum with iodine. Our results are shown in figure 2. We doubt also whether zone 7 of their report is a true lycopene isomer, though we have not been able to free this clearly defined yellow zone from a component which, with iodine, gives a small maximum at 498 mμ.

*Red Tomatoes:* These typically contain phytofluene and all-trans β-carotene. ζ-carotene has not been unequivocally detected in *RRTT* strains. Traces of γ-carotene, and predominantly all-trans lycopene complete the pigment complex.

In all cases, above the lycopene fractions, a yellow-orange zone is discerned. Sometimes this shows xanthophyll characteristics and spectrum. More often the spectrum is ill-defined, and recovery by elution is poor. In no case does this comprise a significant fraction.

*Relative Abundance of Components:* Extracts were prepared, as described above, from composite samples involving several plants of a given line, and by using larger quantities of fruit, up to 5 lb., extracts could be transferred to petroleum ether and made relatively concentrated in pigment without heating in any manner, so that for spectroscopic work, dilutions of 1 to 100 or 1 to 200 were necessary. Such concentrated extracts gave best results for this phase. Aliquots (usually 5 ml.) of the concentrated solution were adsorbed on a column. The absorption spectrum of the properly diluted original solution was immediately determined and an optical density calculated for the concentrated solution (column 2, table 1, etc.). The absorption for the individual components was then determined, in successive columns, and the similarity between the sum of the individual fractions (column 3) and the original (column 2) indicates the extent to which the absorption has been accounted for. This is of particular interest in the tangerine tomato where the absorption indicates the prevalence of pigments of low tinctorial power with high absorption below the main absorption for prolycopene.

Because speed in separation was essential, as many as three adsorbent columns were used, the chromatograms being developed from the start with solvent containing different percentages of acetone (usually 0–2, 5 and 10, though solvent with as much as 15% acetone has been used).

TABLE 1

YELLOW TOMATO: OPTICAL DENSITIES OF EXTRACT AND OF FRACTIONS

| WAVELENGTH | ORIGINAL | SUM OF FRACTIONS | PHYTOFLUENE | β-CAROTENE SET | MISCELLANEOUS* | LYCOPENE* (IMPURE) |
|---|---|---|---|---|---|---|
| 500 | 1.60 | 1.21 | .. | 0.6 | 0.15 | 0.46 |
| 490 | 3.12 | 2.46 | .. | 1.56 | 0.45 | 0.45 |
| 480 | 6.10 | 4.42 | .. | 2.53 | 1.22 | 0.67 |
| 465 | 8.38 | 5.08 | .. | 2.50 | 1.70 | 0.88 |
| 445 | 9.41 | 6.75 | .. | 3.07 | 1.95 | 1.16 |
| 425 | 7.87 | 4.93 | .. | 2.30 | 1.50 | 1.13 |
| 400 | 5.00 | 3.05 | .. | 1.31 | 0.92 | 0.82 |
| 370 | 1.80 | 1.11 | 0.20 | 0.25 | 0.34 | 0.32 |
| 350 | 1.19 | 0.86 | 0.21 | 0.24 | 0.17 | 0.24 |

* The miscellaneous group includes minor components and a larger yellow fraction immediately below the lycopene. Both yellow and lycopene fractions require several chromatograms for spectroscopic purity. A well-defined maximum for the lycopene exists *ca* 505 mμ.

There is a two-fold advantage in reporting the data in terms of optical densities at this stage. It avoids arbitrary compromises in the value selected for the absorption constants. Thus components in table 2 designated $L_1$, $L_2$ and $L_3$ are undoubtedly lycopene-*like* isomers. They have been crystallized, but the curves differ, even after iodine catalysis, and also from the curves obtained prior to crystallization (the change being due to partial isomerization during concentration).

Secondly it enables a comparison to be made between the original optical density, column 2, and the sum of the components, column 3. Losses for more strongly adsorbed components account in part for the differences. This procedure is useful in showing whether there may be a significant contribution by an undetected pigment, and whether unsuspected isomerization has occurred to any extent during evaluation.

The wave-lengths for which optical densities are reported in the tables were chosen on the basis of the characteristics of the absorption curves for the originals, cf. figure 1. Routine measurements were made throughout the whole range, however, without detection of abnormalities.

TABLE 2

TANGERINE TOMATO: OPTICAL DENSITIES OF EXTRACT AND OF FRACTIONS

| WAVE-LENGTH | ORIGINAL | SUM OF FRACTIONS | PHYTOFLUENE | β-SET | ζ-CAROTENE | CIS ψ-CAROTENE | PROLYCOPENE | YELLOW ZONE | $L_3$* | $L_2$ | $L_1$ | $X$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 490 | 30.1 | 25.4 | .. | 3.0 | .. | 0.7 | 11.2 | 1.0 | 1.9 | 2.8 | 4.5 | 0.8 |
| 480 | 46.5 | 41.2 | .. | 5.9 | .. | 1.2 | 18.3 | 1.7 | 3.5 | 3.5 | 5.1 | 2 |
| 465 | 84.8 | 64.1 | .. | 5.9 | <1.0 | 4.8 | 27.5 | 3.9 | 6.2 | 5.0 | 7.0 | 2.8 |
| 435 | 127.5 | 121.3 | .. | 5.8 | 20.6 | 14.2 | 56.5 | 4.2 | 7.0 | 4.8 | 6.0 | 2.2 |
| 425 | 156.3 | 144.4 | .. | 4.3 | 53.1 | 15.6 | 50.0 | 3.3 | 5.5 | 4.1 | 5.2 | 2.3 |
| 410 | 117.5 | 101.2 | .. | 2.9 | 27.1 | 14.1 | 40.4 | 3.1 | 5.0 | 3.3 | 4.3 | 2 |
| 400 | 125 | 119.5 | 0.2 | 2.2 | 53.3 | 13.8 | 35.0 | 2.4 | 4.8 | 2.7 | 4.8 | 1.8 |
| 370 | 68.1 | 61.4 | 12.1 | 1.3 | 21.3 | 5.4 | 15 | ... | 2.0 | 1.8 | 2.5 | ... |
| 360 | 51.3 | 47.2 | 8.8 | 0.5 | 16.0 | 4.0 | 11.6 | ... | 2.0 | 1.8 | 2.5 | ... |
| 350 | 43.8 | 43.5 | 15.6 | 0.4 | 11.9 | 2.3 | 8.6 | ... | 1.6 | 1.4 | 1.7 | ... |

* The designations $L_1$, $L_2$ and $L_3$ represent lycopene isomers separated by a small yellow zone from prolycopene. $X$ is a xanthophyll fraction. $L_3$ contains several minor zones.

TABLE 3

RED TOMATO: OPTICAL DENSITIES OF EXTRACT AND OF FRACTIONS*

| WAVE-LENGTH | ORIGINAL | SUM OF FRACTIONS | PHYTOFLUENE | β-CAROTENE | LYCOPENE |
|---|---|---|---|---|---|
| 500 | 117 | 103.1 | ... | 2.3 | 100.8 |
| 495 | 97.9 | 84.2 | ... | 3.2 | 81.0 |
| 480 | 101.2 | 90.0 | ... | 5.2 | 84.8 |
| 470 | 140.6 | 119.6 | ... | 5.0 | 114.6 |
| 445 | 102.8 | 85.3 | ... | 5.4 | 79.9 |
| 425 | 63.7 | 50.9 | ... | 4.3 | 46.6 |
| 400 | 33.7 | 25.7 | 0.2 | 2.3 | 23.2 |
| 370 | 18.9 | 15.7 | 5.2 | 0.4 | 10.1 |
| 360 | 17.1 | 14.8 | 3.2 | 0.3 | 11.3 |
| 350 | 16.9 | 15.3 | 6.1 | 0.4 | 8.8 |

* γ-Carotene and xanthophyll are minor contributors, *ca* 0.5 at their maxima. The lycopene value is low owing to incomplete elution.

With the reservation that because of the necessity of rapid separation, the zones are not highly purified from a spectroscopic standpoint, the values reported in tables 1, 2 and 3 are a sound measure of the relative abundance of the various components.

The data for tables 1, 2 and 3 are based on equivalent weights of fresh fruit, so that a rough comparison of absolute values is possible. The various yellow lines differ greatly in the amount of pigment produced, and there may be some significant variations with respect to individual components. This phase must be deferred at present. Based on a comparison of table 1 with tables 2 and 3, it appears with respect to all lines of composition *rrTT* that they are markedly deficient in their ability to produce carotenoid pigment; that no component has yet been identified as unique to the yellow tomato; and finally that no component has been found in greater absolute amount in the yellow than in tangerine or in red lines.

When table 2 and table 3 are compared, the total pigment production for tangerine and red lines is approximately equivalent.

As shown in table 1, β-carotene is clearly a major component in the yellow lines and lycopene is present in small amounts. In the red (table 3), the absolute amount of β-carotene is somewhat greater but the proportion, relative to other components, is greatly reduced. The overwhelming contribution is made by all-trans lycopene. The tangerine (table 2) is comparable with the red, so far as β-carotene is concerned, higher in phytofluene, and particularly high in all-trans ζ-carotene, poly-cis ψ-carotene and prolycopene. It will clearly be of interest to isolate the heterozygous genotype *Rrtt* for comparison, with respect to these components.

The evidence indicates certain tentative conclusions as follows:

In the absence of *R*, the *T* gene is responsible for pigment production on a limited scale, particularly with respect to lycopene, cf. reference 2, p. 242. In the absence of *T*, the *R* gene is responsible for large quantities of ζ-, poly-cis ψ-carotenes and prolycopene. In the presence of *R* and *T*, these pigments or their immediate precursors are converted to lycopene. The suggestion of Porter and Zscheile (reference 4, p. 544) that ζ-carotene may be concerned in lycopene formation seems therefore entirely plausible. The open-chain structure recently postulated for ζ-carotene[6] also supports this view.

---

[*] Lines studied in detail include the red "Bonny Best," a tangerine from Dr. J. W. Lesley and the yellows "Snowball" from Good Luck Gardens and MacArthur's tester stock 902.

[1] MacArthur, J. W., *J. of Genetics*, **29**, 123 (1934).

[2] Le Rosen, A. L., Went, F. W., and Zechmeister, L., *Proc. Nat. Acad. Sci.*, **27**, 236 (1941).

[3] Zechmeister, L., Le Rosen, A. L., Went, F. W., and Pauling, L., *Ibid.*, **27**, 468 (1941).

[4] Porter, J. W., and Zscheile, F. P., *Arch. Biochem.*, **10**, 537 (1947).

[5] Zechmeister, L., and Le Rosen, A. L., *J.A.C.S.*, **64**, 1075 (1942).

[6] Nash, H. A., Quackenbush, F. W., and Porter, J. W., *Ibid.*, **70**, 3613 (1948).

# INFRARED ABSORPTION IN FIELD STUDIES OF OLFACTION IN HONEYBEES*

By Walter R. Miles and Lloyd H. Beck

Yale University School of Medicine and Department of Psychology, Yale University

Read before the Academy, November 19, 1947

*Introduction.*—Some two decades ago it was pointed out by Sherrington that the biological function of a specialized sense organ, such as the eye or ear, is to lower the threshold of the nerve fiber to some one form of energy. The senses may be classified in terms of the energy form which serves as the adequate stimulus for each modality. Each sense mechanism that plays a rôle in the integrated behavior of an organism appears subject to stimulation by mechanical energy, chemical energy or radiant energy. According to this widely accepted scheme of classification hearing, touch, pressure, equilibrium sense and the lateral-line organ sense of fishes are all mechano-receptors; vision, the sense of heat or warmth, and the sense of cold are radio-receptors; while the chemo-receptors are smell, taste, common chemical sense and the vomero-nasal organ sense. The theory of olfactory stimulation proposed by the authors[1] would, if substantiated, suggest that the sense of smell be reclassified and placed among the radio-receptors.

In order to test the validity of the hypothesis that stimulation of the olfactory receptors results when the radiation of heat from smell receptors is suddenly increased, a wide range of new investigation is required. Some of the more simply arranged of these desired experimental observations have been undertaken by us, and the present report presents a part of this work, specifically some accomplished with the honeybee, *Apis mellifera* Linn.

Insects are peculiarly well adapted for behavioral studies on the sense of smell due to the peripheral placement of their olfactory receptors along their antennae rather than within a nasal cavity.[2, 3] The worker honeybee offers the additional advantages of high activity level, marked socialization and well-integrated exploratory and work habits. There is a rich scientific literature concerning the olfactory sense in bees. It has been shown that their sense of smell exhibits a marked parallelism with that of man. Von Frisch, as is well known, demonstrated that pairs of different chemical substances judged alike in odor by man were responded to very similarly by the bee. While it is interesting to attempt reinterpretation of older work from the standpoint of a radiation hypothesis, it is obvious that newly designed equipment and methods must be developed and new experimentation conducted in order to obtain critical results for appraisal.

*Apparatus and Methods.*—For our studies on insects we appeared to need what may be called osmic radiation receivers. These may be thought of as artificial noses since they provide chambers which may be filled with odoriferous gases. The apparatus described by us in another paper[1] was designed for some work on cockroaches and is one type of osmic radiation receiver. It is a laboratory type adapted to a variety of insects. A second type of receiver based on the same general principles but considerably simplified has been developed for field studies with bees and other insects. Its general form is illustrated in the accompanying figures 1 and 2. The receiver is made of ordinary black iron pipe fittings as indicated in figure 3. The central part, *B*, is a black iron cross-tee with horizontal openings of one inch, top and bottom openings, three-quarter inch. The inside of this tee serves as a gas chamber and is sealed leak-tight. A black iron diaphragm at *C* closes off one end; *A* is a closed cavity for liquids at the bottom of the gas chamber (we may refer to this as the "perfume bottle"); *F* is a cavity, an extension of the air-tight system of the gas chamber, which is filled with calcium chloride, anhydrous ($CaCl_2$) to absorb water vapor. The absorbent is separated from the gas chamber by a number eight mesh wire screen. *D*, represents the infra-red passing window (described more fully below) which sealed off the front of chamber *B*. Finally *E* in figure 3 represents an accessory chamber for temperature control. The bottom side of this chamber is bounded by the black iron membrane brazed to the iron cross-tee at position *C* in the figure. Ordinarily a small amount of dry ice is placed in receptacle *E* to lower the temperature of the gas chamber below that of the environment. The entire iron unit stands eight inches high, six inches deep, from *D*, as its front, and three inches wide; its weight is about seven pounds. By means of a railing flange it is fastened to a wooden base with two screws. The box was made from old well-weathered chestnut, as exhibited in figure 2. A pair of such receivers was used in our field experiments. One unit was "active" in the sense that infra-red could pass into it. The other unit was used as control. Directly behind the thallium bromide iodide window in the control unit was a glass disc. Thus infra-red rays could not penetrate into the gas chamber of the control unit but were reflected back by the glass.

The window in the front of each osmic radiation receiver was made of a crystal of thallium bromide iodide designated KRS-5.[4] Sections were sawed from the crystal and then polished to a thickness of 2.0 mm. These discs were 50 mm. in diameter.[5] By means of "Quartz Cement"[6] they were sealed into recessed grooves that had been cut in the iron collars at the front of each unit. The groove for the control unit was cut deep enough to receive first a glass disc and then the KRS-5 disc next the glass and in front of it. The thallium bromide iodide was therefore the biological contact surface from the outside for both units.[7]

A flame should not be used against KRS-5 as the fumes from it are very toxic. When polishing this material rubber gloves and a mask should be used.[8] Sections of this crystal take and retain a good polish. This is one of the great advantages it possesses over calcium chloride or potassium bromide crystals both of which require frequent repolishing. The polished KRS-5 surface is not observably toxic to bare fingers or to insects in contact with it.

The second and more important advantage possessed by the KRS-5 crystal is its transmission characteristics for infra-red radiation as revealed in figure 4. Beginning at 0.6 $\mu$ (micron) 600 m$\mu$ (millimicrons) or 6000 A (Angstrom units), which is in the region of the yellow band of the visible spectrum, the transmission rises precepitately to 40 per cent. It increases to about 55 per cent for the extreme visible red, 0.7 $\mu$, and then somewhat more slowly rises in the near infra-red reaching 65 per cent for 1 $\mu$ (10,000 A) and at 1.5 $\mu$ levels off at 70 per cent which is maintained evenly to an extreme of 31–32 $\mu$. The heat radiation in the case of human beings would probably cover the range 7.5 to 15 $\mu$ with a maximum at about 9 to 10 $\mu$. The radiation from bees, judging from the similarity of their odor discriminations to those of human beings, probably covers at least a part of this 7.5–15 $\mu$ range. It is obvious that the KRS-5 provides an amply wide and high transmission range for all such experimental purposes.

A pair of osmic radiation receivers such as described above lends itself to various methods of use in the field. Both windows should have the same temperature which is lower than the environs. (1) The two windows may be exposed side by side for attracting bees directly. (2) The windows may be equipped with short tubes or platforms extending in front of them, bait may be placed here and bees artificially transferred to the two baited areas. This latter procedure has proved the more useful experimental plan and lends itself to the following four variations.

(*a*) Bait removal method: Bees transferred to baited tubes. Before the bait has been completely used up they are gently shaken from the tubes below the units and the bait tubes are removed. Predictions: Bees should return to the window areas. The window that now presents the bait odor (or more of it) should for a time collect the more bees on or near it.

(*b*) Bait exhaustion method: Bees transferred to baited tubes, the tubes are left undisturbed by the experimenter for 20 minutes or more. Predictions: After the bait in the tubes has been exhausted the window that continues to provide more bait odor should hold the attention of more bees, more bees should stay for a time in this tube or bees should more frequently return to this tube.

(*c*) Bait interchange method: Same as (*b*) above except that within one or two minutes after the bait has been exhausted the tubes should be exchanged in position between the units with the least possible disturbance of the few bees in them. Predictions: Bees in the tube placed in front of the experiment window should tend to remain and the bees that have been placed in front of control window should tend to return to the experiment window. If some bees have made trips from and to the bait a temporary position habit may interfere with the predicted result.

(*d*) Bait repetition method: Same as (*b*) above except that when the original bait is nearly exhausted a little more is added equally in each tube and this process is repeated from time to time to keep the bees working for one or more hours, making trips from the bait to the hive and back. Predictions: If the amount of available bait is kept minimal the window that supplies more bait odor should attract the larger population of bees. In work with (*d*) the right-left positions of the units should be exchanged at intervals of 20 or 30 minutes. The units may also be placed one on top of the other and this position changed in a chance order with the right-left positions. Other changes of orientation should be used such as rotation of the units by 180° or 90°.

If we set up our receiver units on a platform near an active hive of bees on some seasonable day when there is not too much natural nectar available, and if we wait long enough some diligent worker bee will find them. She may alight on the window which passes infrared and move slowly, waving her antennae as if near nectar, but she can't get it. She can't carry anything away but an experience. There is no tangible reward for the effort expended. She is unlikely to return to the hive and engage in a dance to attract other bees out to visit this honey smell. Some method of transient baiting to capture the bees' attention was required if experimentation was to proceed without unduly long delays. The method adopted involved two principal features, see figure 5. First, a short plastic tube was provided as an extension from the opening which surrounded each window. Usually three drops of honey were placed in each tube on filter paper or a bit of cloth attached inside and at the bottom of each tube. Second, to a wand, four feet long, filter paper was bound at one end and smeared with honey. This end of the wand was placed near the opening of a hive; after a few minutes the filter paper was well covered with bees. The wand was then lifted, brought back to the units and 10 or 12 bees, an equal number for each unit, were transferred to the bait tubes by gently rolling the wand in such a way as to walk the bees off into the tubes. Each experiment began thus by causing bees to work in the immediate vicinity of the two windows.

*Results.—Experiment I:* The unit containing the "active" or experimental window was on the left side, see figure 6; the control unit with the window that would not pass infra-red was on the right. The thermometer showed 70°F. for outside temperature. The date was September 28, 1947. The amount of natural nectar available was not great. The method employed was a combination of (*a*) and (*d*) above. After the bees had been transferred to the tubes, they cleaned up the available honey rapidly. Drops of honey were added, equally to each tube, until it was found that the bees were working continuously and well. Photo A was taken at 10:30 A. M. At 10:45 (Photo B) the honeybait seemed exhausted in both tubes, only one bee remained in *CW*, those in *EW* were rushing about. At 10:47 both tubes were removed, and placed in an air-tight container. When the tubes were removed, the bees were gently shaken out of them down near the ground below the units. One minute later (Photo C) ten bees were found on *EW*, none on *CW* but three were flying about *CW*. This difference could occur by chance once in about 1000 times. At 10:49$^1/_2$ (Photo D) nine bees were on the "active" window and three flying about it. None were at the control window. At 10:54 (Photo E) three bees were on *EW*, and none on or near *CW*. At 11:09 (Photo F) three were around *EW*, no bees were near *CW*. During the next 15 minutes both windows were continuously observed, but no photographs were taken. Within this period 12 bees alighted on "active" window and eight paused in flight in front of it. No bees alighted on the control window, and only two paused in flight in front of it. This difference represents a chance of one in about 4000.

*Experiment II:* This illustrative experiment follows method (*c*). The tubes were left in place after the bait had diminished practically to the vanishing point judged by the behavior of the bees. Later when most of the bees were found in tube *EW*, the tubes were gently exchanged with the bees in them. This experiment was made in mid-October, 1947. At 9:35 A.M. measured honey bait was placed in the two tubes; no bees came on their own initiative. During the next few minutes bees were transferred equally. After 9:50 A.M. no more bees were transferred. The "active" window (*EW*) was now on the right side; the control unit (*CW*) on the left. At first the bees were working in equal numbers in the two tubes. By 9:57 most of them were in tube *EW;* only one was in front of *CW*. At 9:59 all were in front of *EW*, results to be expected by chance about once in 250 times. At 10:03 most all bees were in tube *EW*. At 10:07 all present were in *EW*. At 10:08 the tubes were gently exchanged and it was not observed that any of the bees left during the exchange. This maneuver brought the great majority of the bees in front of the control window. At 10:20 all were observed to have collected at the "active" (*EW*) window. At 10:21 the tubes were again gently exchanged. At 10:25 some had returned to the "active" window. At 10:34 most of them had returned. By 10:37 all bees present were in the tube in front of the "active" window. A total of 47 minutes had now elapsed since any bees had been transferred to the tubes; 30 minutes since the bait, as a tangible reward, had worn out. These results are typical for other experiments we have conducted.

Field experiments such as the two described above do not, as a rule, yield such definite results when there is abundant nectar to be gathered from wild or cultivated plants available to the bees. This general finding would appear predictable, since the reward the bees secure from their efforts in collaboration with our experiments is meager in comparison with what an equal amount of time devoted to abundant field bloom would give.

Results for experiments attempted in summer when the field supply of nectar was very plentiful lend themselves to presentation in tabular form. These tests were made on a farm in Otsego County, N. Y. in mid-August, 1948.[9] The bees were very busy on buckwheat and red clover, fields of both were lying within a quarter mile of the colonies studied. Method (*b*) was employed, i.e., after bees had been transferred to the baited tubes the situation remained undisturbed for 20 minutes or more. Tests made on the same day succeeded one another at short intervals. Our grouped data for eight tests are presented (see table 1) in terms of the numbers of bees counted at *CW* and at *EW*, 5, 10, 15, and 20 minutes, respectively, after completion of the transfer of the bees to the bait: each test provided four pairs of counts. The 32 pairs for the eight tests show 20 counts in which the number of bees at *EW* exceeded the number at *CW*, and 12 counts in which the count at *CW* equalled or was larger than for *EW*.

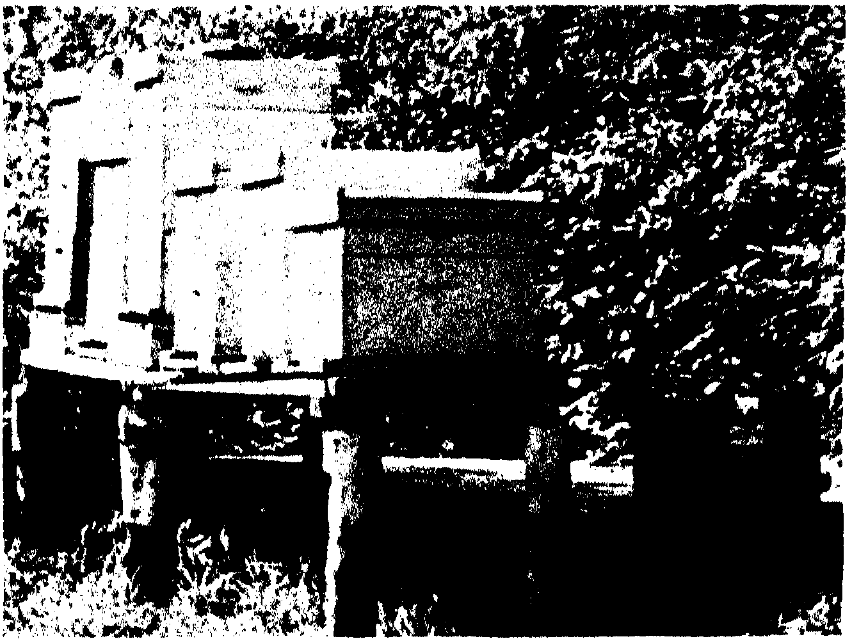
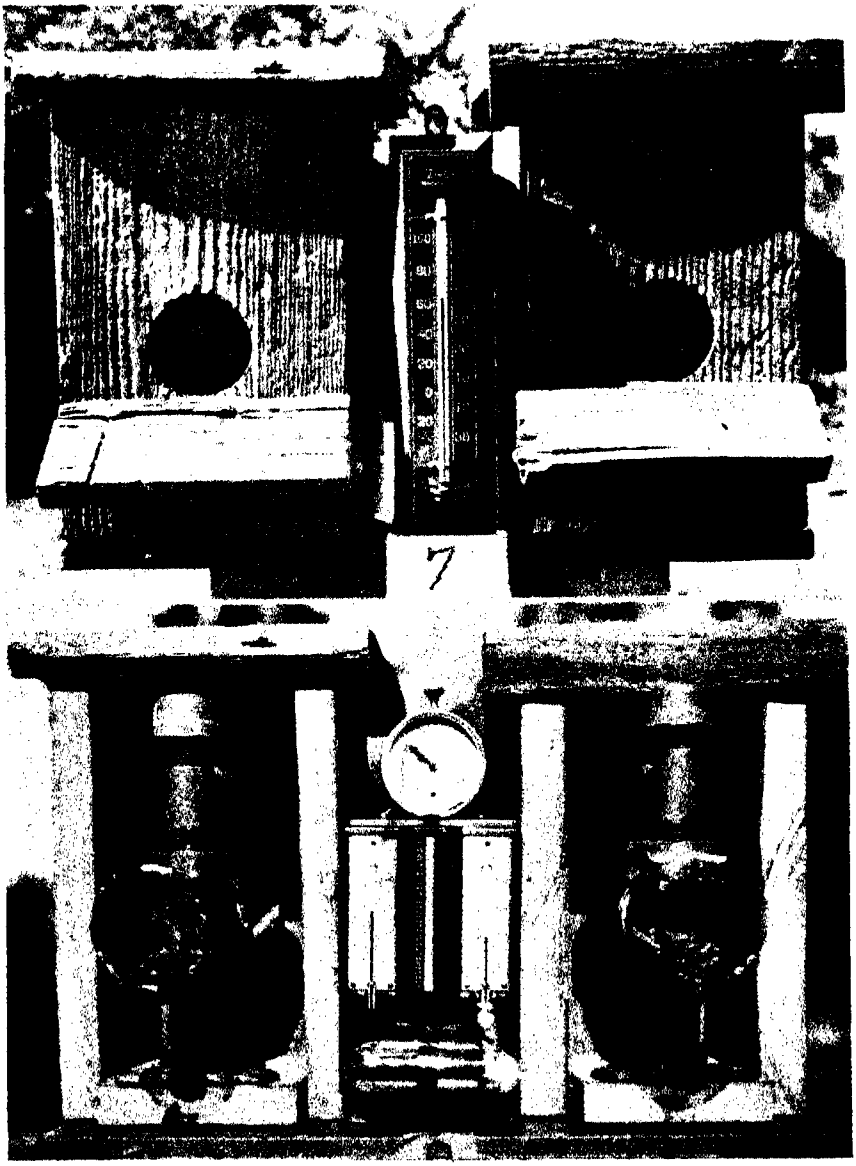


FIGURE 1

A pair of osmic radiation receivers in their housings, placed side by side as used in some olfactory experiments with honeybees, is shown in the lower right-hand corner of the illustration. The two round windows in the fronts of these boxes are the critical areas in reference to which behavior of the bees is observed.

At the bottom of table 1 the totals for the populations found at *CW* and at *EW*, respectively, and for the four time intervals specified have been set down. In the eight tests a total of 133 bees were set at work in front of *CW* and 132 at *EW*. Each of these values was taken as 100 per cent and



on this basis the population percentages remaining at each of the windows after 5, 10, 15 and 20 minutes were computed. In all four pairs of values the percentage retained at *EW* exceeds that for *CW*. The per cent differences are +13.6, +21.6, +7.7, and +6.9, for 5, 10, 15 and 20 minute intervals, respectively, in favor of the experimental window. In summary this

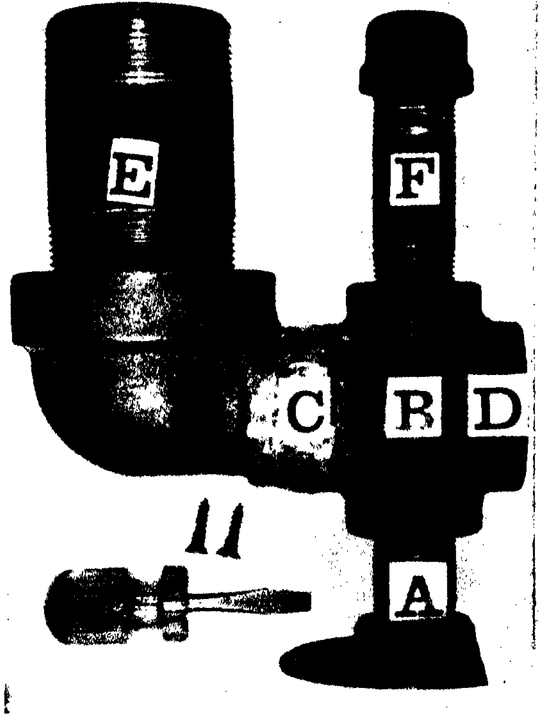


FIGURE 3

Lateral view of osmic radiation receiver which has been formed from black iron pipe fittings. *A*, receptacle for liquid; *B*, gas chamber above surface of liquid; *C*, closed back of chamber formed of black iron diaphragm; *D*, window sealing the front opening of the chamber; *E*, accessory chamber, open at top, for dry ice or other means of temperature control; *F*, an extension of air-tight chamber, *B*, filled with calcium chloride for absorbing water vapor. The unit as illustrated weighs 7 pounds. Scale in cm.

FIGURE 2

The upper view represents a pair of osmic radiation receivers mounted on a tripod-supported platform. The window in the right-hand unit will pass infra-red radiation into the air-tight gas chamber behind it, and is designated the "experimental" window. The control or blank unit is on the left in this view. The window in this unit will not pass infrared radiation into its gas chamber for the reason that a rear segment, made of ordinary glass, reflects such radiation.

The lower view represents the pair of units with the fronts of the housings removed. Plastic tubes 9 cm. in length and 5.5 cm. in diameter are shown fitted to the window collars and projecting horizontally. On the inside and at the bottom of each tube an attached bit of cloth provided a surface on which a small measured amount of honey bait was placed. In this view a number of bees may be seen working in each tube. Ordinarily the fronts of the housing boxes are in place.

TABLE 1

GROUPED DATA FOR EIGHT SUMMERTIME FIELD TESTS OF OLFACTORY RESPONSE IN HONEYBEES

| TEST NO. | DATE, 1948 | HOUR | BEE POPULATION AT START OF TEST | | NUMBER OF BEES COUNTED IN BAIT TUBE | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | AFTER 5 MIN. | | AFTER 10 MIN. | | AFTER 15 MIN. | | AFTER 20 MIN. | |
| | | | *CW* | *EW* | *CW* | *EW* | *CW* | *EW* | *CW* | *EW* | *CW* | *EW* |
| 1 | 8/17 | 5:06 P. M. | 18 | 18 | 16 | 18 | 8 | 14 | 5 | 11 | 2 | 9 |
| 2 | 8/18 | 9:40 A. M. | 22 | 22 | 10 | 22 | 6 | 16 | 5 | 3 | 2 | 0* |
| 3 | 8/18 | 10:05 A. M | 18 | 18 | 16 | 18 | 11 | 15 | 7 | 10 | 4 | 2* |
| 4 | 8/18 | 11:41 A. M. | 18 | 18 | 18 | 18* | 7 | 10 | 4 | 7 | 1 | 3 |
| 5 | 8/19 | 10:35 A. M. | 18 | 18 | 16 | 18 | 11 | 9* | 7 | 2* | 0 | 2 |
| 6 | 8/19 | 12:20 P. M. | 11 | 10 | 10 | 8* | 10 | 7* | 3 | 3* | 0 | 3 |
| 7 | 8/19 | 2:23 P. M. | 12 | 12 | 10 | 11 | 7 | 11 | 0 | 7 | 0 | 4 |
| 8 | 8/19 | 2:55 P. M. | 16 | 16 | 10 | 10* | 6 | 12 | 7 | 5* | 8 | 3* |
| Totals: Pop. honeybees | | | 133 | 132 | 106 | 123 | 66 | 94 | 38 | 48 | 17 | 26 |
| Per cent of pop. after various intervals | | | 100 | 100 | 79.6 | 93.2 | 49.6 | 71.2 | 28.6 | 36.3 | 12.8 | 19.7 |

* Pairs of populations in which *CW* (control window) is as large or larger than *EW* (experimental window).

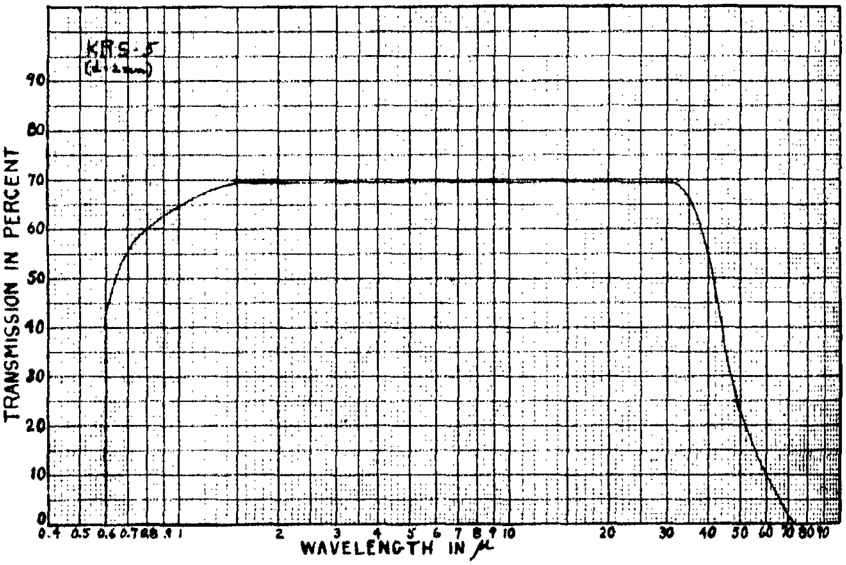


FIGURE 4

Transmission curve for thallium bromide iodide crystal window, 2 mm. in thickness.

group of eight tests, made under conditions when the bees were very actively working at near-by fields of bloom and therefore not easily distracted by artificial stimuli, yields results which tend to support the olfactory radiation hypothesis.

In the foregoing experiments the honey bait was placed inside plastic tubes which extended from the windows. The tubes were an easy fit on the iron collars surrounding the windows and also in the openings in the wooden housings. The tubes were completely open at their front ends. Tubes rather than flat projections were selected as supports for the bait because they would tend to hold the bait odor, locate the insects present in a favorable position for radiating into the windows, and restrain them from crawling about on the wood surrounding the windows. Furthermore, the tubes were easy to manipulate in placement, withdrawal and exchange.

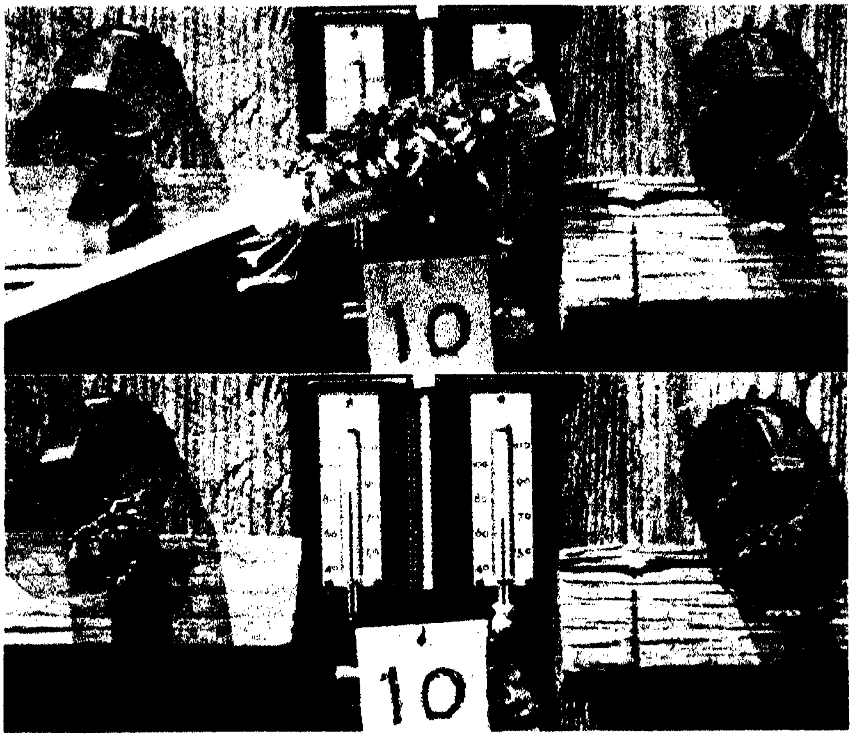


FIGURE 5

Transfer of honeybees to baited areas extending from the windows of the osmic radiation receivers. The upper view shows the transfer wand held in front of the units just before walking the bees off into the tubes. The lower view represents the condition immediately after the bees have been transferred to the baited tubes.

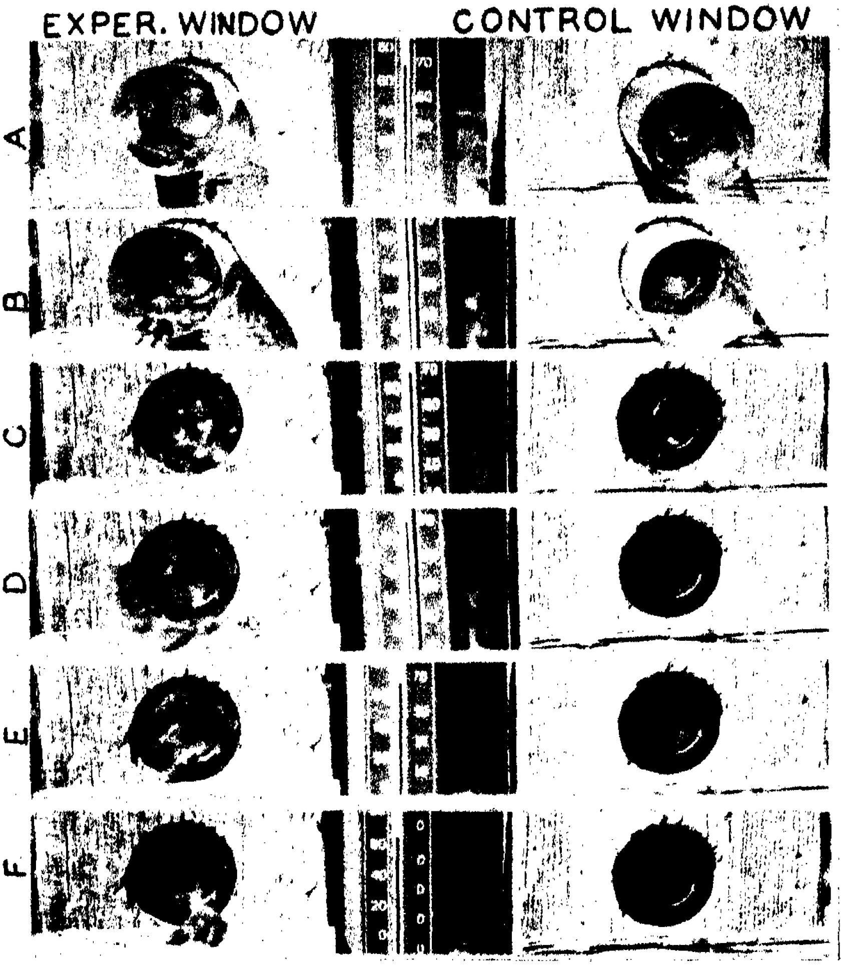


FIGURE 6

A succession of photographic views which from top to bottom constitute the record of one experimental test. *A* shows the bees immediately after transfer to the bait areas; *B* represents the condition when the bait is about exhausted. Between *B* and *C* the bait tubes have been gently removed and the bees shaken out of them near the ground directly below the two units. *C* and the other views, to the bottom of the figure, illustrate at succeeding intervals of time the predominating tendency of the bees to gather on or near the experimental window, which is on the left in this illustration.

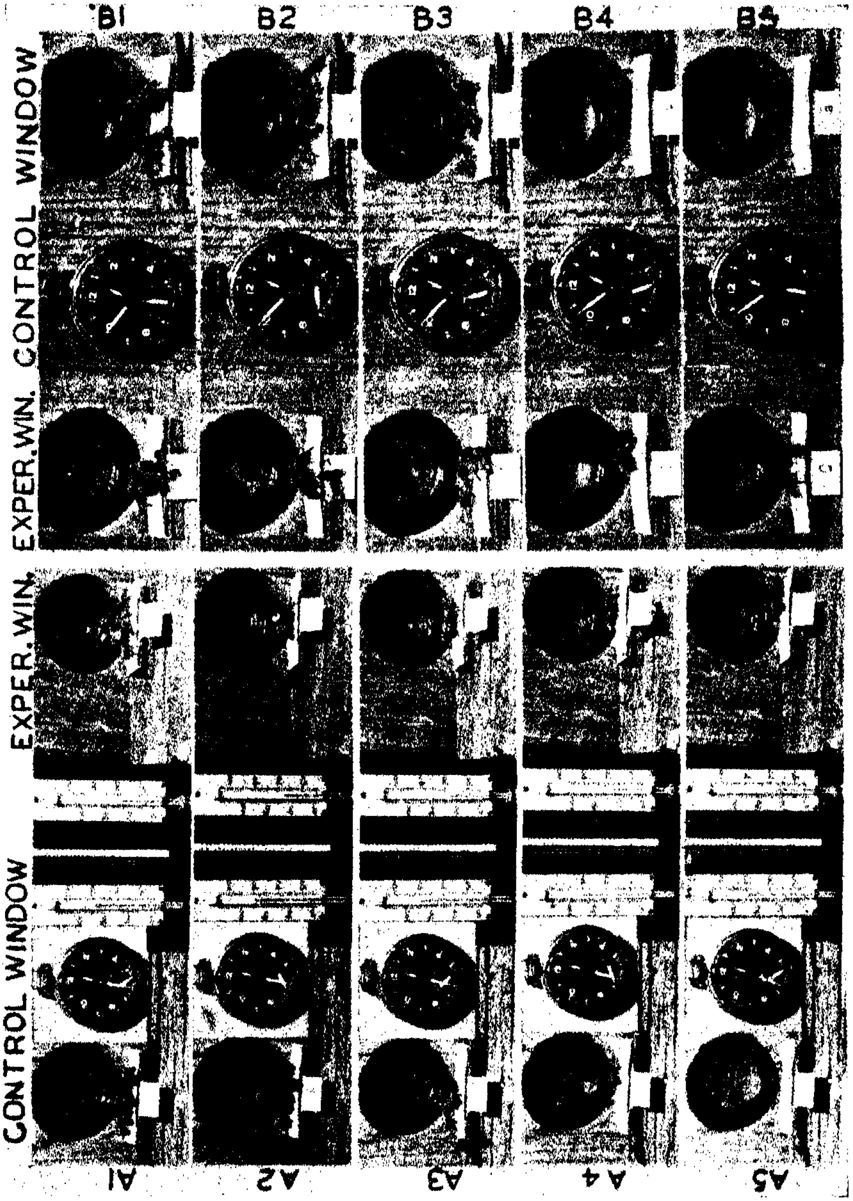


FIGURE 7

Two sets of photographic records representing two succeeding tests in which the experimental window is at the right in the first, and at the left in the second test. In place of tubes for holding the bait, small platforms have been arranged. For purposes of this illustration the photographs of the timepiece which during the experimentation were above the thermometers and near the tops of the unit housings, have been mounted in line with the windows to facilitate compactness of illustration. For discussion of the results of these two tests see the text.

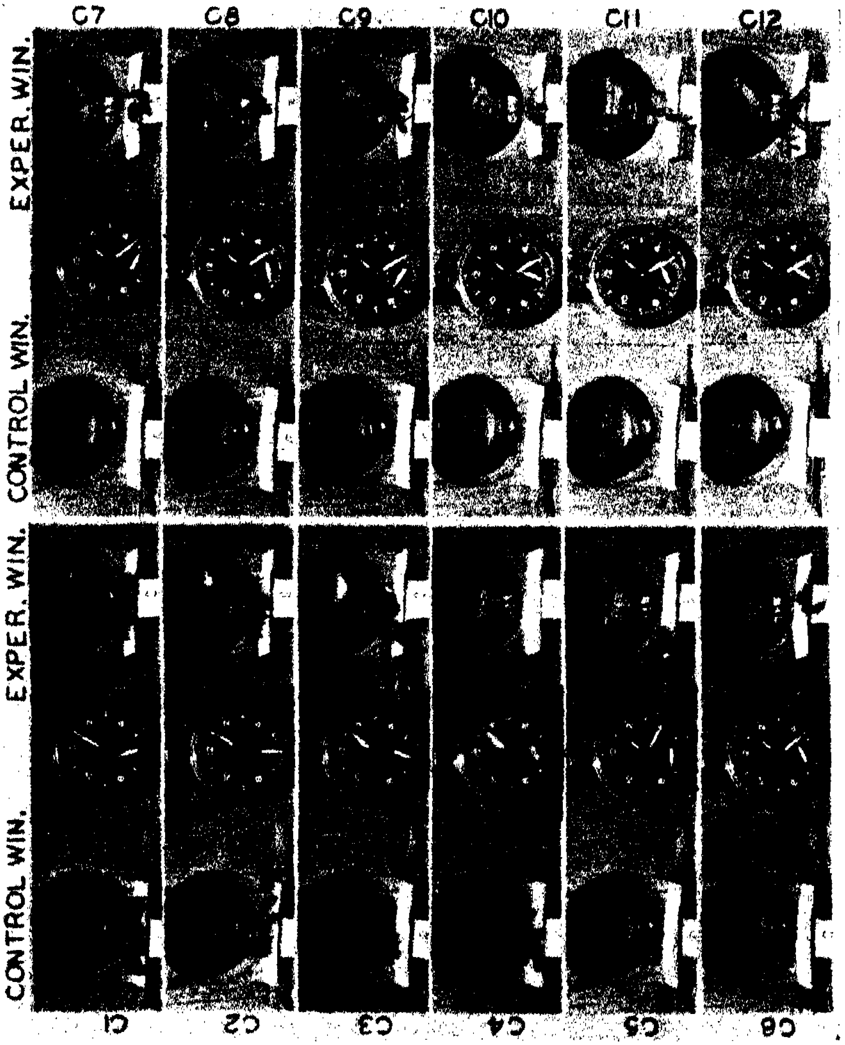


FIGURE 8

The 12 individual photographs composing this illustration constitute the record for one osmic radiation test. The beginning of the test is represented by the extreme upper left-hand view, C1, the end of the test by the bottom right-hand view, C12. When the test began platform *C* was below the experimental window. Immediately after the third view in the sequence, taken at 12:06, the two platforms were exchanged, as far as possible without disturbing the bees.

However, they have one disadvantage for the experimenter, since he finds them something of a visual obstruction, when attempting to count the bees present, and also inconvenient when attempting to record the populations photographically.

Early in November, 1948, a few experiments were tried using flat, plastic platforms attached at the bottom of each window collar and extending to the front and horizontally from that position. These platforms were two inches wide and one inch deep. By means of adhesive tape, a small amount of white cotton gauze was attached and covered the mid-portion of each portion. A series of three tests of this type are here reported. These were made at the same Connecticut farm at which our work had been done the previous fall.[10] In these tests a special effort was made to secure good photographic recording of the populations present at varying time intervals.[11] The season of the year was favorable; wild nectar was scarce. But we predicted that with the flat open platforms the bees would remain around the windows for a shorter time than when bait was placed in the tubular supports.

Test A was made shortly after 12:30 P.M.; the results are photographically illustrated in the left-hand portion of figure 7. In preparing this figure the original photographs have been cut down so as to include principally the two windows. The photos of the face of the timepiece have been brought down from their original position near the tops of the housings, and pasted between the windows for the sake of compactness in the illustration. The two platforms in these tests were lettered *W* for the control window and *C* (crystal) for the experimental window. The control window (*W*) was on the left, the experimental window on the right. Three drops of honey were carefully placed near the center of each platform, and eight worker bees were transferred to each platform. Photo A1 was taken immediately after the bees had been transferred and had started working. Fifteen seconds later, A2, eight bees still working in front of each window. Photo A3 made 26 seconds after A1, still shows most of the bees present on each side. In photo A4, made one minute and 16 seconds following A1, a single bee is on *W*, in front of the control unit, while three bees are on *C*, in front of the experimental window and two more are near about. The final photo, A5, taken two minutes after A1, reveals a characteristic result, there are no bees on platform *W* while five are shown on platform *C*.

Test B followed shortly after Test A and the five photographic records of it are reproduced in the right-hand portion of figure 7. It will be noted that the right-left positions of the two units have been exchanged, and in Test B the experimental window is now on the left-hand side. Again three drops of honey were placed centrally on each platform and eight worker bees were transferred to each platform. Photo B1 was made just following this transfer. Photo B2 was made 43 seconds later. The bees had practically exhausted the available bait, and were now scattered and moving about. B3, taken 57 seconds after B1, indicates that some of the insects had left each platform. In Photo B4, recorded two minutes and 26 seconds after B1, four insects may be seen on platform *C*, near the experimental window, none on platform *W*. The final photo, B5, made three minutes after the beginning, records three bees on platform *C* and none on *W*. The reader will note that both of these tests, A and B, followed method (*b*) described as "the bait exhaustion method."

Test C, after a short interval, followed the previous two tests, and was made according to method (*c*) in which the bait platforms were exchanged in the early portion of the test. This experiment extended for 21 minutes, 12 photographic records were taken and all have been reproduced in figure 8. The two units had again been exchanged in their right-left position. Three drops of honey were placed on each platform and ten worker bees were transferred to each. Photo C1 records the conditions present immediately after the two equal populations had been set to work. It should be stated that two or three bees were transferred at a time to each platform, by moving the wand first to one and then to the other, in turn, so that neither platform would suffer the handicap that the bees got to work on it earlier than on the other one. In Photo C2, 34 seconds after C1, the bees on platform *W* had begun to move about and one or two had left. In Photo C3, 39 seconds after C1, nine or ten bees may be seen on platform *C*, only three on platform *W*. Shortly after this view was made no bees remained on platform *W*. Then the experimenter exchanged the positions of the two platforms without greatly disturbing the bees resting on platform *C*. As he completed the exchange, he unfortunately jostled the tripod of the camera so that the picture taken, Photo C4, was somewhat blurred. Nevertheless, it indicates that the bees were successfully transferred on platform *C* to the position in front of the control window. The bees soon began leaving platform *C;* this is not recorded in the picture sequence. For about 13 minutes no more than one bee at a time would appear on platform *W*. If another came, he would be fought off. Some would fly about the experimental window briefly. Photo C5, made nearly 13 minutes after Photo C4, and 20 minutes after C1, records two insects at platform *W*, and none on platform *C*. In fact, from 1:20 P.M., C5, until the end of the test, 1:26 P.M., C12, no bee alighted on platform *C* whereas on *W* by the experimental window the population ranged from two to five. The bees on *W* tended to fight each other and to fight off other bees that attempted to join them in their frustrated search for the honey which they seemed to sense was present.

*Discussion.*—The experiments on honeybees reported in this paper have constituted an attempt on our part to put the olfactory radiation hypothesis to test. These field trials followed some laboratory tests on cockroaches carried out with an apparatus built according to the same principles embodied in the osmic radiation receivers here described. Our initial trial with bees was on September 1, 1947, at an apiary on a farm in Madison, Connecticut. In this test the beekeeper kindly coöperated with us. He introduced a small frame smeared with honey into the top of one of the hives. The smell of honey within the hive caused the bees to become active, in a search for honey, and within 15 minutes there were large numbers of them outside the hive on some frames from which the honeycombs had been removed, but small bits naturally still remained. We now considered that the bees had been sufficiently stimulated to find honey in their near vicinity. At first we placed our two units on a suitable tripod support, facing the sun, and with their backs about 20 feet away from the hive. Watching the units for 30 minutes, no bees were observed to pay any attention to either of the windows. We then changed the position of the units so as to face the hive and moved them about seven feet away from it. Again observing for 30 minutes, we noted that no bees came to either window. We then placed the units on the ground directly in front of the hive entrance and three feet away from it. The dry ice in the cooling chambers of each box had caused the iron window frames to frost up and moisture to gather on the windows. The moisture was removed and we observed the units for 15 minutes; no bees seemed to pay any attention to either box or window. The units were moved still nearer to the hive, and again no positive response. It was now 5:30 P.M.; we moved the boxes to within nine or ten inches from the mouth of the hive. Boards immediately below and in front of the hive formed a suitable platform for the units. There was activity at the mouth of the hive, workers were coming and going, also some workers were stinging drones and pushing them to one side. The two units were about one inch apart, and blocked low-angle approach of the bees to the hive opening. Under these conditions and after 15 minutes of watching, we observed no approach by the bees to the boxes themselves, that is, no bee lighted on either box, but one or two bees did pause a moment in flight in front of the experimental window. At this point we placed a thin string around both boxes in such a way that its position horizontally bisected each window. By means of a penknife blade, a small drop of honey was placed on the string directly in front of each window. After about five minutes there appeared a fly, then a yellowjacket. The yellowjacket perched on the string in front of the experimental window and consumed the drop of honey, then explored up and down the string immediately in front of the window. Presently he found the honey in front of the control window, where the fly had been engaged. The fly was frightened off and the yellowjacket consumed that drop of honey also. The yellowjacket now returned to the experimental window, reexamined the string where the honey had been, then flew up in front of the window. He returned to the string in front of the control window, examined the string, but did not pay any obvious attention to the window. During the next 15 minutes a yellowjacket (apparently the same one) returned frequently, paused in front of the experimental window or lighted on the string or window or on the frame close to it. So far as we could judge, he paid about the same amount of attention to the string in front of the window as he did to the window itself. His behavior in reference to the experimental window was much more positive than in reference to the control window, in front of which he had gotten an equal amount of honey, and was last rewarded. These initial trials on our first day of experimenting led us to devise the technique of supplementary baiting in short plastic tubes projecting from the window areas.

It seems probable that the bait-tube technique can be improved. The flat platforms were considerably narrower than the length of the tubes. The front edges of the platforms extended only $1^{11}/_{16}$ inches or 43 mm. from the vertical planes of the front surfaces of the thallium windows. Tubes of this effective length or even shorter might be used and the bait surface disposed around an arc of the inside circumference rather than lengthwise in the tube. Transferred bees would then tend to have an initial orientation facing inward toward the bait and also toward the windows. Such refinements together with a technique for marking the bees set to work at the window areas might yield more quantitative and more analytical results than we have thus far secured.

In the tests reported above we have varied the position and distance of the receiver units from the hive. The two units alternated in their respective right-left positions. The temperatures of the two units were kept nearly equal, and below the environmental temperature but not always at the same differential between environment and the units. Different times of day in mid-summer and in late fall have been tried. The bait tubes were used interchangeably, that is, one of them was not made to be used at the experimental window, and the other at the control window. The tubes were sometimes (method *c*) exchanged after the beginning of a test, sometimes in other tests (method *a*) they were removed after the experiment was well started. The windows were often wiped dry if they tended to gather moisture, and they were kept clean. Honey from the same colony worked with was always used as the bait. Throughout all these controls and variations of testing the bait-seeking behavior of bees introduced into the situation seems much more positive in reference to the infrared passing window. It seems fair to assume that in front and near this window the insects have an olfactory experience and that it is this which causes them to collect and to remain for some minutes. After the bait has become exhausted the situation, for those that remain, would seem to offer no primary reward. These bees sensing honey, perhaps strongly, but not finding it seem irritable. At least it is characteristic of their behavior that in the latter half of a test they fight off other comers that approach the experimental window. Except for this behavioral factor the population differential between the windows might have been more striking. A technique of working with a group of bees, one at a time, as reported by Frings,[13] should avoid any such masking effect introduced by this element of social interaction. Our field experiments unfortunately lacked this type of control.

If it is fair to assume that the bees collect at *EW* because of an olfactory experience gotten there which is not found at *CW*, then on the physical side it may be assumed that the behavioral difference is based on a difference in infrared radiation. The experiments indicate that the difference is due to radiation lost by the insects, in all probability from their antennae.[3, 12, 13] The experiments reported here do not settle the question whether the radiation is direct to the cooled gas molecules of honey vapor in the air-tight chamber beyond the thallium window, or whether ambient honey-vapor molecules adsorbed on the antennae provide a filter-like film through which radiation takes place, and is accentuated by the cooled chamber behind the transmitting window near the insects. In the latter case it would probably make little difference whether the chamber was filled with honey gas or some other gas. On the basis of some other experimentation not here reported, we are inclined to believe that the second alternative is the one that is operating. In either case a radiation hypothesis, as the stimulus for olfaction in honeybees, seems to be supported.

*Conclusions.*—1. It is suggested that the initial physical event in the stimulation of the olfactory receptors is due to an increase in their heat radiation caused by the heat absorption of certain vapors or gases.

2. Osmic radiation receivers, designed to provide infrared passing windows for air-tight gas chambers, are described.

3. A technique and methodology of using osmic radiation receivers in field studies on honeybees is outlined.

4. Experimental tests on the behavior of honeybees in reference to the osmic radiation receivers are described and critically evaluated. Controls were introduced for a number of variables.

5. In general, the results support the hypothesis. More bees definitely paid attention and attended for longer intervals to the infrared passing window than to the window of the control unit. The experiments support the interpretation that the behavioral difference is due to increased heat losses from the insects' antennae.

---

* This investigation was supported in part by grants from Fluid Research Funds, Yale University, and from the Research Corporation of New York City.

[1] Beck, L. H., and Miles, W. R., "Some Theoretical and Experimental Relationships between Infrared Absorption and Olfaction," *Science*, **106**, (A) 511 (1947).

[2] Wigglesworth, V. B., *The Principles of Insect Physiology*, E. P. Dutton & Co., New York (1939), pp. 147–152. Snodgrass, R. E., "The Morphology of Insect Organs and the Sensory Nervous Systems," *Smithsonian Inst. Pub. Misc. Collections*, **77** (8), 1 (1926).

[3] McCord, C. P., and Witheridge, W. N., *Odors: Physiology and Control*, McGraw-Hill Book Co., New York (1949). See extensive bibliography.

[4] Three brief abstracts concerning KRS-5 (composed of mixed crystals of thallium bromide and thallium iodide) appear in *J. Opt. Soc. Am.*, **38**, 663–664 (1948).

[5] The KRS-5 discs were produced at Engineer Research and Development Laboratories, The Engineer Center, Fort Belvoir, Va. C. R. Hoover, Project Engineer, Radiation Engineering Branch, supplied the transmission chart for KRS-5 (thickness of 2.0 mm.) and stated in a letter dated July 3, 1947, that the "transmission characteristics of KRS-5 are no longer classified. . . ." Paul J. Ovrebo of Wright Field, through his kindness, was instrumental in helping us secure the use of two discs of this recently perfected and expensive material.

[6] Product of Lakeside Chemical Co., So. Chicago Ave., Chicago. Mr. Beardsley of Wright Field kindly gave us the benefit of his experience in the use of this material.

[7] The KRS-5 disc in the control unit developed an observable crack while being sealed in its collar against the glass. We had no alternate disc to substitute and therefore sealed the crack. This seal across the face of the disc may be noted in some of the reproduced photographs.

[8] Cooke, F., "Optical Finishing of KRS Materials," *J. Opt. Soc. Am.*, **37**, (A) 948 (1947). High polish can be obtained through the use of pitch and beeswax laps.

[9] We wish to acknowledge the kind coöperation of Frederick A. Godley, owner and operator of this farm in New York.

[10] We wish to acknowledge the kind coöperation of William C. Peck, owner and operator of this Connecticut farm where we have conducted most of our tests on honey-bees.

[11] *Life* Magazine photographer Eric Schaal was happily with us on this occasion. We have permission to use some of the photographs which he took of our experiments. See figures 7 and 8.

[12] Frings, H., "The Loci of Olfactory End-Organs in the Honey-Bee, *Apis mellifera* Linn," *J. Exper. Zool.*, **97**, 123–134 (1944).

[13] Crozier, W. J., "Chemoreception," Ch. 19, pp. 987–1036, *Handbook of General Experimental Psychology*, Murchison, C., Editor, Clark Univ. Press, Worcester, Mass. (1934).

[14] Beck, L. H., "Osmics: Theory and Problems Related to the Initial Events in Olfaction," *Medical Physics*, Glasser, Otto, Editor. Second edition in press.

# THE BEADED MINUTE COMBINATION AND SEX-DETERMINATION IN DROSOPHILA

BY A. H. STURTEVANT

KERCKHOFF BIOLOGICAL LABORATORIES, CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA, CALIFORNIA

Communicated March 19, 1949

In a recent paper[1] Goldschmidt reports that Beaded Minute males of *Drosophila melanogaster* are usually intersexual, the effect being produced by any one of 10 out of 12 different Minutes tested. From experiments with these types he draws certain conclusions concerning the mechanism of sex-determination in Drosophila.

I have examined Beaded Minute flies in which eight different Minutes were concerned (including three of those from which Goldschmidt reports more extreme effects), and have found numerous aberrant males in seven of these combinations; but detailed study has convinced me that these are not intersexes. They represent cases in which the external genitalia are imperfectly developed or completely absent, but there is no clear indication of femaleness, either in the external genitalia or in any of the internal parts. The male genital ducts are often rudimentary, but neither they nor the gonads show any trace of femaleness. In normal males and females there is a sexual difference in the anal plates (two lateral ones in the male, a dorsal and a ventral in the female), and it has been shown[2, 3] that one of the most sensitive indices of intersexuality is the presence of three plates (two laterals and a ventral). I have carefully examined over 350 Beaded Minute males, from many cultures, with different Minutes and different genetic backgrounds, that were grown at widely different temperatures; in no case have I found any ventral anal plate. It is, however, frequently the case that the whole anal and genital area is smooth, with no sign of anal or genital opening or sclerite. Such individuals, when dissected, show very rudimentary genital ducts (usually not found at all, but perhaps overlooked), and a gut that is closed behind the rectal glands, its blind end lying free in the body cavity. These males, without exception, have had two normal sex-combs, and normal male-type abdominal tergites. The published accounts of triploid intersexes do not indicate any tendency to abortion of the anal plates, and I have recently examined over 200 triploid intersexes of varying grades without finding any case in which less than two anal plates were present.

The above account was submitted to Dr. Goldschmidt, who has kindly given me his comments on it. (See the paper immediately following this one.) He agrees with many of the observations, but has made additional ones which convince him that these specimens are really intersexes. Some of these observations concern the shapes of the parts in the more extreme specimens, and leave me unconvinced. Dr. Goldschmidt has, however, called my attention to a few hairs that are often present on the seventh sternite (sixth in the terminology of Newby[3]) of Beaded Minute males, in positions very similar to those regularly found on the seventh sternite of normal females. I had overlooked these hairs, but can now confirm the observation that they are often present. However, I find that they are occasionally present in normal (and fertile) males of a wide variety of constitutions. They were even present in an *X Y* male that was homozygous for the gene *tra*,[6] and that therefore can scarcely be considered as in any degree intersexual. These hairs are, then, not good indications of intersexuality. Their frequent presence in Beaded Minute males is interesting from a developmental point of view, but is not to be taken as a sign of femaleness.

We are then dealing not with intersexuality but with incomplete development of the anal and genital imaginal discs—a condition that is well known in a wide variety of Drosophila types as an occasional anomaly, and is a regular characteristic of certain genotypes—e.g., *In* (*2*) *ho* (Lewis[4]). In agreement with this conclusion is the fact that Beaded Minute *females* also often lack anal plates and external anal opening, though this happens less frequently than in Beaded Minute males. In five Beaded Minute females the genital opening was also absent. The tendency to imperfect development of imaginal discs, in Beaded Minute of both sexes, is most marked at the posterior end, but is still much greater than is usually the case in other parts of the body. Among the defects rather frequently seen may be mentioned roughened eyes with reduced area, mis-shapen legs, incompletely unfolded wings, and incompletely chitinized abdominal tergites.

Goldschmidt argues that the primary effect here is that of Beaded, the Minutes serving to accentuate it. My observations do not support this view, since I have seen such aberrant specimens in rather casual examinations of seven out of fifteen Minute strains with no Beaded ancestry. This result suggested that the effect is part of the "Minute reaction" studied by Schultz.[5] In accord with this conclusion is the fact that Hairless largely suppresses the effect, as it does the "Minute reaction." Delta accentuates the "Minute reaction," and accordingly Delta-3 was used. Delta-3 with Minute-3w or with Minute-S13 does in fact give a series of males and females showing the same anal and genital defects as does Beaded Minute; but so also does Delta-3 Beaded, and I have found an exactly similar series in a wholly unrelated culture in which only a few second chromosome mutant genes (not Minutes) were involved. In all these cases the series of types found is very closely similar to those in Beaded Minute, and the frequencies are well within the range of variation among strongly affected Beaded Minute cultures. It is clear, therefore, that this type of defect is rather easily produced, by a wide variety of genetic combinations.

Goldschmidt suggests that the Beaded section of the third chromosome may turn out to be one of the two male determiners of Drosophila. This suggestion rests on the two conclusions that I have here suggested are incorrect—namely, that the Beaded Minute aberrant males are intersexual, and that their peculiarities are due primarily to the Beaded locus.

*Notes on Dr. Goldschmidt's Paper That Immediately Follows This One.*—On many points Goldschmidt and I are in agreement. It seems clear that the reduction of the anal plates in both sexes is not in itself relevant to the question of sex determination. I should also conclude that reduction of the male genital apparatus, and the appearance of hairs on the seventh sternite, are indicative of intersexuality only if they can be shown to be associated with other, more unambiguous, female characteristics. In my material they were not so associated; Goldschmidt finds them to be so, but only occasionally—as he suggests, perhaps owing to genetic differences between the strains used. Not having found Beaded Minute males with such properties, I can only conclude that they arise as a consequence of some other, unanalyzed, genetic composition—and that discussion of their relation to general theories of the genetic determination of sex is premature.

*Summary.*—Beaded Minute males are not intersexual, but are often aberrant because of incomplete development of the anal and genital imaginal discs—as are also Beaded Minute females.

Even if they are considered to be intersexual, the abnormality may be induced by any one of at least seven different Minutes, or by other independent combinations of genes, and therefore cannot be used to support the view that there are very few loci involved in the determination of sex.

[1] Goldschmidt, R. B., PROC. NAT. ACAD. SCI., **34**, 245–252 (1948).

[2] Dobzhansky, T., and Bridges, C. B., *Amer. Nat.*, **62**, 425–434 (1928).

[3] Newby, W. W., *Univ. Texas Pub.* 4228, 113–145 (1942).

[4] Lewis, E. B., *Genetics*, **30**, 137–166 (1945).

[5] Schultz, J., *Ibid.*, **14**, 366–419 (1929).

[6] Sturtevant, A. H., *Ibid.*, **30**, 297–299 (1945).

# THE INTERSEXUAL MALES OF THE BEADED MINUTE COMBINATION IN DROSOPHILA MELANOGASTER

By Richard B. Goldschmidt

University of California

Communicated March 26, 1949

Dr. A. H. Sturtevant kindly sent me a copy of the manuscript of his paper on the Beaded Minute males of *Drosophila melanogaster*. He holds that these males described by me as intersexual males are not real intersexes, but products of rudimentation of the genital and anal discs. There is no doubt that many individuals might be described this way if they stood alone. But this is also true for many triploid intersexes of Dobzhansky's classes I–III (Dobzhansky, 1930). Actually, these types are only one in a series of morphological changes which indicate clearly that we are dealing with male intersexes. It seems that, by chance, Sturtevant met only with the one type which usually, though not always, is the most frequent one. Detailed statistics on the frequency and genetics of the different types will be published in time together with the necessary illustrations of all the types, grades and transitions.

It should be stated first that the lower grade Bd-M intersexes are different from the triploid intersexes (Dobzhansky's classes I, II). Both the anal plates and the type of reduction of the armature are very different. Sturtevant is much impressed by this fact. But it should be pointed out that, thus far, the only male intersexes available for comparison are the triploid ones. For female intersexes a number of different strains are known, all of which have their specific features, which show that the details of intersexual transformation vary with the different genetic backgrounds. Thus, I cannot regard the triploid intersexes as an absolute standard of intersexual structure. But it is a fact that the internal structure of the Bd-M intersexes is exactly the same as that described by Dobzhansky for the triploids of his classes I–III. Unfortunately, the higher classes of intersexuality are rather rare in the Bd-M case.

As Sturtevant mentions only the one type of intersexuality which I call the "reduction type" it should be proved first that this type is also actually an expression of intersexuality in which the reduction and final disappearance of the male genital armature is more prominent than the secondary appearance of female characters. This type is a combination of two phenomena, namely, the reduction and disappearance of the anal plates and that of the genital armature proper. It is the conspicuous reduction of the anal plates found in this type only, but absent in the other types, which led Sturtevant astray. I think that his statements and interpretations are correct as far as the anal plates are concerned. In a rather good correlation with the intensity of the Bd-M effect upon the males, the anal plates and even the abdominal tip are reduced in a part of the females, and some not Beaded-Minutes may show this also. But it will be shown below that the involution of the anal plates is independent of the reduction of the genital armature. In the reduction type, now under discussion, the genital armature is retracted and found (if not completely lost) below the surface, in conditions which indicate that it had developed to a certain degree—all degrees from completeness to total absence—and then stopped growing. (This includes also stoppage at different times of the well-known 360° rotation in the pupa.) These conditions parallel closely those described by Dobzhansky for the triploid intersexes, although not in every detail. The highest grade of this reduction type is characterized by a rather large abdomen without any genital structure at the tip. As obviously the entire disc region degenerated early, no further female differentiation is possible. But in less extreme grades (called in my work class III) the disc is still present below the integument, and in this stage I twice found (in not too many dissections of this specific type, as yet) individuals with rudimentary testes, without any male ducts, but with a complete vagina with its characteristic histological differentiation, and with a ventral receptacle complete in one, not perfect in the other case. In the higher grades of this type, hairs are found rather frequently on the 7th sternite. This is a female character. Examination of innumerable normal brothers of the same genetic constitution never showed this phenomenon.

A second type of intersex appears rather irregularly in the $F_1$ and $F_2$ broods. In our earlier work it was just as frequent as the reduction type in certain Bd-M combinations. Recently it was found rather rarely, indicating that specific modifiers may control these types. (All types may appear in the same brood. Statistics will be published in time.) In this second type there is no degeneration of the anal plates. Instead, they even hypertrophy. They grow out beyond the tip of the abdomen into two large plates which frequently unite dorsally (a female trait). The dorsal part of the genital arch to which the plates are attached is also pushed out as a socket, which, if sufficiently large, develops hairs of the type of the female 8th tergite, which it clearly represents. The genital armature proper undergoes the same stages of reduction as described for the first type until the ventral side of the abdomen below the large beak-like protruding anal plates is free of differentiations. In still higher grades this ventral surface becomes depressed and forms a deep cup. In more extreme individuals, the anterior lip of the cup-like depression grows backward until only a curved slit remains. In the only three dissections of this type made thus far the bottom of the covered cup was found to have the histological structure of the vagina (uterus) and is therefore considered as the secondary invagination of a vagina. (I agree with those authors who think the vaginal disc is not identical with the male disc.)

A third type, which we call the direct type is rather frequent in its lower grades, but thus far rather rare in the higher grades. It represents a simple and direct transformation of the male abdominal tip into an almost female one. There are, of course, also transitions and combinations between the three main types, e.g., degeneration of the anal plates with progressive differentiation of the female segment. The only male parts to become rudimentary in this type are the claspers and the penis, which assume very characteristic shapes during this process. The first step is the fusion of the two ventral tips of the genital arch with their characteristic tufts of hair into a single median one, which protrudes as a knob. The genital arch then broadens and becomes oriented as an 8th segment. The possibility that a 9th segment is involved will be discussed another time. The knob becomes larger and is transformed into a protruding structure—still with the tufts of hair—resembling the female abdominal tip and assumes the spatial relation to the anal plates typical of the female. Finally the genital arch forms a typical female eighth (or 8th and 9th?) segment with its normal pigmentation and arrangement of hairs. Also, the tuft of hairs on the "knob" scatters over the sternite. In one such extreme type the anterior surface of the knob carried a disc which might have been a vaginal disc or the anlage of the vaginal plates. Only rarely are the anal plates arranged in a dorsal and ventral position, as in females.

An accurate record of these facts requires many illustrations which will be published when all dissections are completed. It should be added that in these experiments thousands of brothers of the intersexes of other genetic constitutions, i.e., Bd or M or neither, were constantly watched as controls without a single case of intersexuality or reduction of the genital armature occurring. There is no doubt in my mind that the types described are male intersexes, which leaves the theoretical considerations which I made in my earlier paper (Goldschmidt, 1948) and in another one now in press (Goldschmidt, 1949) as they were originally formulated.

Dobzhansky, Th., *Bull. Bur. Genetics* (Leningrad), **8**, 91–158 (1930).
Goldschmidt, R. B., PROC. NAT. ACAD. SCI., Wash., **34**, 245–252 (1948).
Goldschmidt, R. B., *Arch. Julius Klaus-Stiftung*, **23**, 1948: 539–549.

# A THEORY OF TRACE-ADMISSIBLE ALGEBRAS

By A. A. Albert

Department of Mathematics, The University of Chicago

Communicated April 30, 1949

1. *Introduction.*—There have been a number[1] of studies of non-associative algebras in which the lack of the associative law has been compensated for by the assumption of one or more other identities. The principal tool in most of these studies has been a trace argument,[2] and the simple algebras, which are the end results, have invariably been of relatively special types.

In the present note we shall make a major advance in the theory of non-associative algebras by obtaining a general theory which includes all previous theories in which the radical was defined to be the maximal nilideal. We shall restrict our attention to power associative algebras over a field $\mathfrak{F}$ whose characteristic is prime to thirty, and shall call such and algebra $\mathfrak{A}$ *trace-admissible* if there is a bilinear function $\tau(x, y)$ with arguments $x$ and $y$ in $\mathfrak{A}$ and values in $\mathfrak{F}$ such that

$$\tau(x, y) = \tau(y, x), \qquad \tau(x, yz) = \tau(xy, z), \tag{1}$$

$\tau(e, e) \neq 0$ if $e$ is an idempotent of $\mathfrak{A}$, $\tau(x, y) = 0$ if $xy$ is a nilpotent element of $\mathfrak{A}$ or is zero. We define the *radical* of $\mathfrak{A}$ to be its maximal nilideal[3] $\mathfrak{N}$, and say that $\mathfrak{A}$ is *semisimple* if $\mathfrak{N} = 0$. We also say that $\mathfrak{A}$ is *simple* if $\mathfrak{A}$ is semisimple and $\mathfrak{A}$ and 0 are the only ideals of $\mathfrak{A}$.

If $\mathfrak{A}$ is any algebra there is an attached commutative algebra $\mathfrak{A}^{(+)}$. This is the same vector space as $\mathfrak{A}$ but is an algebra defined by a product $x \cdot y$ expressible in terms of the product $xy$ of $\mathfrak{A}$ by the formula $2x \cdot y = xy + yx$. We shall show that when $\mathfrak{A}$ is trace-admissible then $\mathfrak{A}$ and $\mathfrak{A}^{(+)}$ have the same radical. Moreover, when $\mathfrak{A}$ is semisimple, the trace function $\tau(x, y)$ of $\mathfrak{A}$ is an admissible trace function for $\mathfrak{A}^{(+)}$. But then we can show that *$\mathfrak{A}^{(+)}$ is a Jordan algebra, $\mathfrak{A}$ is flexible, $\mathfrak{A}$ is simple if and only if $\mathfrak{A}^{(+)}$ is simple.* It follows that *the only simple trace-admissible algebras* are the (commutative) *Jordan algebras,* the *flexible* (noncommutative) *algebras of degree two,* and the (noncommutative flexible) *quasiassociative algebras.* These results show that unless some technique other than a trace argument is utilized in studying power associative algebras no new simple algebras will arise.

2. *Elementary Properties.*—If $e$ is an idempotent of a power associative algebra $\mathfrak{A}$ over $\mathfrak{F}$ of characteristic prime to 30 it is known[4] that $\mathfrak{A}$ can be decomposed as the supplementary sum

$$\mathfrak{A} = \mathfrak{A}_e(1) + \mathfrak{A}_e({}^1/_2) + \mathfrak{A}_e(0) \tag{2}$$

of subspaces $\mathfrak{A}_e(\lambda)$ such that

$$ea_\lambda + a_\lambda e = 2\lambda a_\lambda \tag{3}$$

for every $a_\lambda$ of $\mathfrak{A}_e(\lambda)$. Then it is actually true that $ea_1 = a_1e = a_1$, $ea_0 = a_0e = 0$. The subspaces $\mathfrak{A}_e(1)$ and $\mathfrak{A}_e(0)$ are subalgebras of $\mathfrak{A}^{(+)}$, and this implies that $a_\lambda b_\lambda + b_\lambda a_\lambda = c_\lambda$ is in $\mathfrak{A}_e(\lambda)$ for every $a_\lambda$ and $b_\lambda$ of $\mathfrak{A}_e(\lambda)$ with $\lambda = 0, 1$. It is also true that

$$a_\lambda b_{1/2} + b_{1/2}a_\lambda = c_{1/2} + c_{1-\lambda} \tag{4}$$

where $c_{1/2}$ is in $\mathfrak{A}_e(^1/_2)$ and $c_{1-\lambda}$ is in $\mathfrak{A}_e(1 - \lambda)$ for every $a_\lambda$ of $\mathfrak{A}_e(\lambda)$ and $b_{1/2}$ of $\mathfrak{A}_e(^1/_2)$ with $\lambda = 0, 1$. Finally

$$a_{1/2}b_{1/2} + b_{1/2}a_{1/2} = c_0 + c_1, \tag{5}$$

where the notations should be evident at this point.

The subspaces $\mathfrak{A}_e(1)$ and $\mathfrak{A}_e(0)$ are not, in general, subalgebras of $\mathfrak{A}$. However, since they are subalgebras of $\mathfrak{A}^{(+)}$, and powers in $\mathfrak{A}$ and in $\mathfrak{A}^{(+)}$ coincide, it is true that the polynomial algebra $\mathfrak{F}[a_\lambda]$ is contained in $\mathfrak{A}_e(\lambda)$ for every $a_\lambda$ of $\mathfrak{A}_e(\lambda)$, $\lambda = 0.1$.

The results about the decomposition above are consequences, in the case where $\mathfrak{A}$ is commutative, of the identity $x^2x^2 = (x^2x)x$. If we replace $x$ by $x + \phi y$ in this identity the coefficient of $\phi$ must be zero and this yields the basic identity[5]

$$yx^3 = 4(yx)x^2 - (yx^2)x - 2[(yx)x]x. \tag{6}$$

We shall use this identity in the proof of an essential lemma.

LEMMA 1. *Let $\mathfrak{B}$ be an ideal of a power associative algebra $\mathfrak{A}$ and $e$ be an idempotent of $\mathfrak{A}$ such that $\mathfrak{A}_e(0)$ is contained in $\mathfrak{B}$. Then $g^3$ is in $\mathfrak{B}$ for every $g$ of $\mathfrak{A}_e(^1/_2)$.*

For $\mathfrak{B}$ is also an ideal of $\mathfrak{A}^{(+)}$, and $g^3$ is the same in both $\mathfrak{A}$ and $\mathfrak{A}^{(+)}$. We may then assume, without loss of generality, that $\mathfrak{A}$ is commutative. We use (5) to write $g^2 = c_0 + c_1$ and put $x = g$, $y = e$ in (6) to obtain $e[(c_0 + c_1)g] = 2g^3 - c_1g - g^3 = (c_0 + c_1)g - c_1g = c_0g$. Since $c_0$ is in the ideal $\mathfrak{B}$ so are $e(c_0g)$ and $c_0g$. Then $e(c_1g)$ is in $\mathfrak{B}$. But $c_1g = d_{1/2} + d_0$ by (4), $e(c_1g) = {}^1/_2 d_{1/2}$ is in $\mathfrak{B}$, $d_0$ is in $\mathfrak{B}$, $c_1g$ is in $\mathfrak{B}$, $(c_1 + c_0)g = g^3$ is in $\mathfrak{B}$ as desired.

3. *Trace Admissible Algebras.*—Let $\mathfrak{A}$ be power associative, $e$ be an idempotent of $\mathfrak{A}$, $\tau(x, y)$ be an admissible trace function for $\mathfrak{A}$. Use (2) and note that

$$\tau(x_1, y_0) = \tau(x_1, y_{1/2}) = \tau(x_0, y_{1/2}) = 0 \tag{7}$$

for every $x_\lambda$ of $\mathfrak{A}_e(\lambda)$ and $y_\mu$ of $\mathfrak{A}_e(\mu)$. Indeed $x_1y_0 = 0$, $\tau(x_\lambda, y_{1/2}) = \tau(x_\lambda, ey_{1/2} + y_{1/2}e) = \tau(ex_\lambda + x_\lambda e, y_{1/2}) = 2\lambda\tau(x_\lambda, y_{1/2})$ and so, if $2\lambda \neq 1$, we have $\tau(x_\lambda, y_{1/2}) = 0$. Observe that we have used the consequence

$$\tau(x, yz) = \tau(zx, y) \tag{8}$$

of (1).

Any admissible trace function for $\mathfrak{A}$ is trivially an admissible trace function for any subalgebra $\mathfrak{B}$ of $\mathfrak{A}$. It is known[6] that the *radical* (maximal ideal of nilpotent elements) of $\mathfrak{A}$ *is the set* $\mathfrak{N}$ *of all elements* $x$ *of* $\mathfrak{A}$ *such that* $\tau(x, y) = 0$ *for every* $y$ *of* $\mathfrak{A}$. Thus $\mathfrak{N}$ is independent of $\tau$.

An idempotent $e$ of $\mathfrak{A}$ is said to be *principal* if there exists no idempotent $u$ orthogonal to $e$. Since every element of $\mathfrak{A}_e(0)$ is orthogonal to $e$ the space $\mathfrak{A}_e(0)$ contains no idempotent. But if $a$ were a non-nilpotent element of $\mathfrak{A}_e(0)$ the algebra $\mathfrak{F}[a]$ is associative and would contain an idempotent $w$ which would be in $\mathfrak{A}_e(0)$. Hence all elements of $\mathfrak{A}_e(0)$ are nilpotent when $e$ is principal. We may then prove[7]

LEMMA 2. *Let* $e$ *be a principal idempotent of a trace admissible algebra* $\mathfrak{A}$ *with radical* $\mathfrak{N}$. *Then* $\mathfrak{A}_e(^1/_2) + \mathfrak{A}_e(0)$ *is a subspace of* $\mathfrak{N}$.

For the sum $x_0 + y_0$ of any $x_0$ and $y_0$ of $\mathfrak{A}_e(0)$ is nilpotent, so is $(x_0 + y_0)^2$, $0 = \tau(x_0 + y_0, x_0 + y_0) = \tau(x_0, x_0) + \tau(y_0, y_0) + 2\tau(x_0, y_0)$. But $x_0^2$ and $y_0^2$ are nilpotent, $\tau(x_0, x_0) = \tau(y_0, y_0) = 0$, $\tau(x_0, y_0) = 0$. By (7) $\tau(x_0, y) = 0$ for every $y$ of $\mathfrak{A}$, $\mathfrak{A}_e(0)$ is a subspace of $\mathfrak{N}$. By Lemma 1 we see that $g^2$ is in $\mathfrak{N}$ for every $g$ of $\mathfrak{A}_e(^1/_2)$, $g$ is nilpotent. Then $\tau(x_{1/2} + y_{1/2}, x_{1/2} + y_{1/2}) = 2\tau(x_{1/2}, y_{1/2}) + \tau(x_{1/2}, x_{1/2}) + \tau(y_{1/2}, y_{1/2}) = 0$, $\tau(x_{1/2}, y_{1/2}) = 0$ for every $x_{1/2}$ and $y_{1/2}$ of $\mathfrak{A}_e(^1/_2)$, $\tau(x_{1/2}, y) = 0$ for every $y$ of $\mathfrak{A}$ by (7), $\mathfrak{A}_e(^1/_2)$ is contained in $\mathfrak{N}$.

4. *Reduction to the Semisimple Case.*—If $\mathfrak{A}$ has a unity quantity 1 we may define $\tau(x) = \tau(1, x)$ so that $\tau(x)$ is a *linear* function of $x$. Then $\tau(x, y) = \tau(1, xy) = \tau(xy)$ by (1) and the bilinear function $\tau(x, y)$ is the usual linear trace function $\tau$ of a product. Note that $\tau(e, e) = \tau(e) \neq 0$ for any idempotent $e$, $\tau(y) = \tau(1, y) = 0$ for $y$ any zero or nilpotent quantity, $\tau(xy) = \tau(yx)$, $\tau[x(yz)] = \tau[(xy)z]$. We now prove

LEMMA 3. *If* $\tau(x, y)$ *is an admissible trace function for an algebra* $\mathfrak{A}$ *with a unity quantity then* $\tau(x, y)$ *is an admissible trace function for* $\mathfrak{A}^{(+)}$ *so that the radicals of* $\mathfrak{A}$ *and of* $\mathfrak{A}^{(+)}$ *coincide.*

For $\tau(x, y) = \tau(y, x)$ is a property only of the vector space $\mathfrak{A}$, $\tau(x, y \cdot z) = {}^1/_2\tau(x, yz + zy) = {}^1/_2\tau(xy + yx, z) = \tau(x \cdot y, z)$ by (8) and (1). If $e$ is an idempotent of $\mathfrak{A}^{(+)}$ then $e$ is also an idempotent of $\mathfrak{A}$ and so $\tau(e, e) \neq 0$. Assume now that $x \cdot y = {}^1/_2(xy + yx)$ is nilpotent. Then $\tau(x \cdot y) = 0$ as was shown above, ${}^1/_2\tau(xy + yx) = \tau(xy) = \tau(x, y) = 0$ as desired. Note that actually $\tau(x \cdot y) = \tau(xy)$ for every $x$ and $y$ of $\mathfrak{A}$.

We shall next prove

THEOREM 1. *Let* $\mathfrak{A}$ *be a trace-admissible power associative algebra over a field* $\mathfrak{F}$ *of characteristic prime to 30, and* $\mathfrak{N}$ *be the radical of* $\mathfrak{A}$. *Then the semisimple algebra* $\mathfrak{A} - \mathfrak{N}$ *is trace-admissible and has a unity quantity so that* $(\mathfrak{A} - \mathfrak{N})^{(+)} = \mathfrak{A}^{(+)} - \mathfrak{N}^{(+)}$ *is semisimple,* $\mathfrak{N}^{(+)}$ *is the radical of* $\mathfrak{A}^{(+)}$.

For the elements of $\mathfrak{A} - \mathfrak{N}$ are the images $x_0$ of the elements $x$ of $\mathfrak{A}$ under a homomorphism $x \to x_0$ of $\mathfrak{A}$ onto $\mathfrak{A} - \mathfrak{N}$. Define $\tau(x_0, y_0) = \tau(x, y)$. Then if also $x_1 \to x_0$ and $y_1 \to y_0$ we know that $x_1 = x + x'$, $y_1 = y + y'$ where $x'$ and $y'$ are in $\mathfrak{N}$, $\tau(x_1, y_1) = \tau(x, y) + \tau(x', y) + \tau(x, y') + \tau(x', y') = \tau(x, y)$ since $x'y$, $xy'$ and $x'y'$ are in $\mathfrak{N}$. The properties (1) are trivial for the function $\tau(x_0, y_0)$ since they hold for $\tau(x, y)$, and if $x_0y_0$ is nilpotent then a power of $xy$ must be in $\mathfrak{N}$, $xy$ is nilpotent, $\tau(x, y) = \tau(x_0, y_0) = 0$. Every idempotent $e_0$ of $\mathfrak{A} - \mathfrak{N}$, in the power associative case, is the image of an idempotent $e$ of $\mathfrak{A}$ and $\tau(e_0, e_0) = \tau(e, e) \neq 0$, $\tau(x_0, y_0)$ is an admissible trace function for $\mathfrak{A} - \mathfrak{N}$.

The algebra $\mathfrak{A} - \mathfrak{N}$ is necessarily semisimple and is now trace-admissible. Let $e$ be a principal idempotent of $\mathfrak{C} = \mathfrak{A} - \mathfrak{N}$. By Lemma 2 we see that $\mathfrak{C}_e(^1/_2) + \mathfrak{C}_e(0) = 0$, $e$ is the unity quantity of $\mathfrak{C}$. By Lemma 3 the algebra $\mathfrak{C}^{(+)}$ is semisimple. But every ideal $\mathfrak{B}$ of $\mathfrak{A}$ determines an ideal $\mathfrak{B}^{(+)}$ of $\mathfrak{A}^{(+)}$, $\mathfrak{A} - \mathfrak{B}$ and $\mathfrak{A}^{(+)} - \mathfrak{B}^{(+)}$ are the same vector spaces. Hence $(\mathfrak{A} - \mathfrak{N})^{(+)} = \mathfrak{A}^{(+)} - \mathfrak{N}^{(+)}$, $\mathfrak{N}^{(+)}$ must be the radical of $\mathfrak{A}^{(+)}$.

5. *The Principal Results.*—The results just derived show that *if $\mathfrak{A}$ is a semisimple trace-admissible algebra then $\mathfrak{A}^{(+)}$ is semisimple and trace admissible.* It must then be true that the first step in the study of the structure of semisimple trace-admissible algebras is the case where $\mathfrak{A}^{(+)} = \mathfrak{A}$, that is, the commutative case. Let us then assume that $\mathfrak{A}$ is commutative so that we can use (6).

The mapping $a \to ax = aR_x$ is a linear transformation $R_x$ on $\mathfrak{A}$ and it will be convenient to write $yR_x^3$ rather than $[(yx)x]x$. We note that for a commutative algebra $\tau(xy, z) = \tau(x, yz) = \tau(xz, y)$ and so

$$\tau(yR_x^k, z) = \tau(y, zR_x^k) \tag{9}$$

for every positive integer $k$. We now have $w = yx^3 - 4(yx)x^2 + (yx^2)x + 2yR_x^3 = 0$, $zw = 0$, $\tau(z, w) = \tau(zy, x^3) - 4\tau(zx^2, yx) + \tau(zx, yx^2) + 2\tau(z, yx^3) = 0$. But by (9) and the commutative law we see that $\tau(zy, x^3) + 2\tau(z, yx^3) = 4\tau(zx^2, yx) - \tau(zx, yx^2)$ is unaltered by the interchange of $y$ and $z$. Thus $4\tau(zx^2, yx) - \tau(zx, yx^2) = 4\tau(yx^2, zx) - \tau(yx, zx^2)$, $5\tau(zx^2, yx) = 5\tau(zx, yx^2)$. If the characteristic of $\mathfrak{F}$ is prime to 30 we have $\tau(zx^2, yx) = \tau[z, x^2(yx)] = \tau(zx, yx^2) = \tau[z, (x^2y)x]$, $\tau[z, x^2(yx) - (x^2y)x] = 0$ for every $z$ of $\mathfrak{A}$, $x^2(yx) - (x^2y)x$ is in the radical $\mathfrak{N} = 0$ of $\mathfrak{A}$, $x^2(yx) = (x^2y)x$. We have proved the first part of the

PRINCIPAL THEOREM. *Let $\mathfrak{A}$ be a semisimple trace-admissible power associative algebra over a field $\mathfrak{F}$ whose characteristic is prime to 30. Then $\mathfrak{A}^{(+)}$ is a semisimple Jordan algebra, $\mathfrak{A}$ is flexible, $\mathfrak{A}$ is uniquely expressible as a direct sum $\mathfrak{A} = \mathfrak{A}_1 \oplus \ldots \oplus \mathfrak{A}_t$ of simple algebras $\mathfrak{A}_t$ such that $\mathfrak{A}_t^{(+)}$ is a simple Jordan algebra. Every simple $\mathfrak{A}$ of order $n$ over a center $\mathfrak{F}$ of characteristic 0 or $p > n$ is either a simple Jordan algebra, or has degree two, or is quasiassociative,*[8] $\mathfrak{A}^{(+)}$ is simple.

To prove the flexibility of $\mathfrak{A}$ we use the consequence

$$w = (xy + yx)x - x(xy + yx) + x^2y - yx^2 = 0 \qquad (10)$$

of the hypothesis $xx^2 = x^2x$. Then $zw = 0$ for every $z$ of $\mathfrak{A}$, and we use (1) and (8) to obtain

$$\tau(xz - zx, xy + yx) = \tau(zy - yz, x^2).$$

Interchange $z$ and $y$ to obtain $\tau(xy - yx, xz + zx) = \tau(yz - zy, x^2) = \tau(xy + yx, zx - xz)$. Add $\tau(xy + yx, xz + zx)$ to both members of the resulting equation and obtain $\tau(xy, xz + zx) = \tau(xy + yx, zx)$. Then $\tau(xy, xz) + \tau(xy, zx) = \tau(xy, zx) + \tau(yx, zx)$, $\tau(xy, xz) = \tau(yx, zx)$, $\tau[x(yx) - (xy)x, z] = 0$, $(yx)x - x(xy)$ is in the radical of $\mathfrak{A}$ and $\mathfrak{A}$ is flexible.

To complete our results we note the known[9]

Lemma 4. *Let* $\mathfrak{B}$ *be an ideal of a trace-admissible algebra* $\mathfrak{A}$. *Then the radical of* $\mathfrak{B}$ *is the intersection of* $\mathfrak{B}$ *and the radical of* $\mathfrak{A}$.

We now prove

Lemma 5. *Let* $\mathfrak{B}$ *be a vector subspace of a semisimple trace-admissible algebra* $\mathfrak{A}$ *and* $\mathfrak{B}$ *be an ideal of* $\mathfrak{A}^{(+)}$. *Then* $\mathfrak{B}$ *has a unity element* $e$, $\mathfrak{B} = \mathfrak{A}_e(1)$ *is an ideal of* $\mathfrak{A}$, $\mathfrak{A} = \mathfrak{B} \oplus \mathfrak{A}_e(0)$, $\mathfrak{A}^{(+)} = \mathfrak{B} \oplus \mathfrak{A}_e(0)$.

For when $\mathfrak{A}$ is semisimple so is $\mathfrak{A}^{(+)}$. By Lemma 2 we know that $\mathfrak{B} = \mathfrak{B}_e(1)$ has a unity quantity $e$. Using products in $\mathfrak{A}^{(+)}$ we see that $2x \cdot e = x$ is in $\mathfrak{B}$ for every $x$ of $\mathfrak{A}_e(^1/_2)$, $x \cdot e = x = 0$, $\mathfrak{B}$ is a subspace of $\mathfrak{A}_e(1)$, $x \cdot e = x$ for every $x$ of $\mathfrak{A}_e(1)$, $\mathfrak{A}_e(1) = \mathfrak{B}$, $\mathfrak{A} = \mathfrak{A}_e(1) + \mathfrak{A}_e(0)$. Since $\mathfrak{A}$ is flexible $\mathfrak{A}_e(1)$ and $\mathfrak{A}_e(0)$ are *orthogonal* subalgebras of $\mathfrak{A}$ (and of $\mathfrak{A}^{(+)}$) and the result is proved.

The result above implies that $\mathfrak{A}$ *is simple if and only if* $\mathfrak{A}^{(+)}$ *is a simple Jordan algebra*. The remaining results of our principal theorem are trivial consequences of this result and a known[10] theorem on the structure of a flexible algebra $\mathfrak{A}$ such that $\mathfrak{A}^{(+)}$ is a simple Jordan algebra.

---

[1] Such algebras are the alternative algebras, right alternative algebras, Lie algebras, Jordan algebras and what the author has called standard and static algebras.

[2] A trace argument can be used in the study of alternative and associative algebras and is the only available tool for the study of Lie algebras. For its use in other cases see the following papers of the author: I. A Structure Theory for Jordan Algebras, *Ann. Math.*, **48**, 546–567 (1947); II. Power-Associative Rings, *Trans. Am. Math. Soc.*, **64**, 552–593 (1948); III. On Right Alternative Algebras, *Ann. Math.*, **50**, April number (1949).

[3] This assumption about the radical cuts out the theory of Lie algebras since all Lie algebras are nilalgebras.

[4] See II in reference 2, pages 558–561.

[5] *Ibid.*, (14) on page 556.

[6] *Ibid.*, page 564.

[7] This result is proved in II (reference 2) under an assumption which cannot be used here.

[8] Quasiassociative algebras are defined on page 583 of II (reference 2).

[9] See Theorem 2 on page 565 of II in reference 2.

[10] *Ibid.*, Theorem 13 on page 593.

---

# THE HOMOTOPY GROUPS OF A TRIAD

By A. L. Blakers and William S. Massey

Department of Mathematics, Princeton University

Communicated by S. Lefschetz, April 2, 1949

In this paper we define new homotopy groups for topological spaces. These groups generalize the homotopy groups of Hurewicz. By the use of these groups and by improved methods we obtain new results about the ordinary homotopy groups, and also easier proofs of known results. Among other things, we can show that $\pi_6(S^3)$ is non-trivial.

1. One of the principal problems of modern topology is to devise methods for computing the homotopy groups of a space. The homotopy groups of even such simple spaces as spheres have not been computed except in special cases. This contrasts strongly with the situation for the homology groups, which can be computed for any triangulable space. The homotopy groups resemble homology groups in all basic properties except one: the homology groups are invariant under an excision,[1] while this is not generally true for the homotopy groups. More precisely, if a space is a union $A \cup B$ of two subspaces, then under fairly general conditions the inclusion map $i:(A, A \cap B) \rightarrow (A \cup B, B)$ (called an *excision*) induces isomorphisms of the corresponding relative homology groups, in all dimensions, but will not generally do so for the relative homotopy groups. This is perhaps the chief reason for the difficulty of computing the homotopy groups. The new homotopy groups defined in this paper are a measure of the deviation from invariance under excision for the relative homotopy groups.

We shall use the following notation for certain subsets of cartesian $n$-space, $C^n$. The coördinates of a point $x \in C^n$ are denoted by $(x_1, \ldots, x_n)$, and $|x| = (x_1^2 + \ldots + x_n^2)^{1/2}$.

$$\begin{aligned}
E^n &= \{x \in C^n \mid |x| \leq 1\},\\
E^{n-1}_- &= \{x \in E^n \mid |x| = 1, x_n \leq 0\},\\
E^{n-1}_+ &= \{x \in E^n \mid |x| = 1, x_n \geq 0\},\\
S^{n-1} &= E^{n-1}_- \cup E^{n-1}_+, \qquad S^{n-2} = E^{n-1}_- \cap E^{n-1}_+,\\
E^n_1 &= \{x \in E^n \mid x_2 \geq 0\}, \qquad E^n_2 = \{x \in E^n \mid x_2 \leq 0\},\\
p_0 &= (1, 0, \ldots, 0).
\end{aligned}$$

Let $(X; A, B)$ be a *triad;* that is, $A$ and $B$ are subspaces of the topological space $X$, and $A \cap B \neq 0$. Choose a base point $x_0 \in A \cap B$. A *map*

$$f:(E^n; E^{n-1}_+, E^{n-1}_-, p_0) \rightarrow (X; A, B, x_0)$$

is a continuous function $f:E^n \rightarrow X$ such that $f(E^n_+) \subset A$, $f(E^n_-) \subset B$, and $f(p_0) = x_0$. Two such maps $f_0$, $f_1$, are *homotopic* if they are connected by a continuous 1-parameter family of maps,

$$f_t:(E^n; E^{n-1}_+, E^{n-1}_-, p_0) \rightarrow (X; A, B, x_0),$$

where $0 \leq t \leq 1$.

LEMMA. *If* $n \geq 3$, *and* $f:(E^n; E^{n-1}_+, E^{n-1}_-, p_0) \rightarrow (X; A, B, x_0)$, *then there exist homotopic maps* $f'$, $f''$, *with* $f'(E^n_1) = f''(E^n_2) = x_0$.

Denote the set of all homotopy classes of mappings $(E^n; E^{n-1}_+, E^{n-1}_-, p_0) \rightarrow (X; A, B, x_0)$ by $\pi_n(X; A, B, x_0)$. For $n \geq 3$, we define an addition between any two elements $\alpha, \beta \in \pi_n(X; A, B, x_0)$ as follows: Choose maps $f$, $g$ belonging to the homotopy classes $\alpha$, $\beta$, respectively, with $f(E^n_2) = g(E^n_1) = x_0$. Define $h:(E^n; E^{n-1}_+, E^{n-1}_-, p_0) \rightarrow (X; A, B, x_0)$ by $h|E^n_1 = f|E^n_1$ and $h|E^n_2 = g|E^n_2$. Then $\alpha + \beta$ is defined to be the homotopy class of $h$. With this definition $\pi_n(X; A, B, x_0)$ becomes a group, called the $n$th *homotopy group of the triad* $(X; A, B)$ *at the base point* $x_0$.

Just as in the case of the relative homotopy groups, a continuous map $f:(X; A, B, x_0) \rightarrow (X'; A', B', x_0')$ induces homomorphisms of the triad homotopy groups. We denote these by $f_*:\pi_n(X; A, B, x_0) \rightarrow \pi_n(X'; A', B', x_0')$.

Associated with the triad $(X; A, B)$ are two *boundary homomorphisms,*

$$\partial:\pi_n(X; A, B, x_0) \rightarrow \pi_{n-1}(A, A \cap B, x_0),$$
$$\partial':\pi_n(X; A, B, x_0) \rightarrow \pi_{n-1}(B, A \cap B, x_0),$$

defined as follows. If $f:(E^n; E^{n-1}_+, E^{n-1}_-, p_0) \rightarrow (X; A, B, x_0)$ represents $\alpha \in \pi_n(X; A, B, x_0)$, $(n \geq 3)$, then $f|E^{n-1}_+$ represents $\partial(\alpha)$, and $f|E^{n-1}_-$ represents $\partial'(\alpha)$. Now consider the following sequence of groups and homomorphisms:

$$\ldots \xrightarrow{\partial} \pi_n(A, A \cap B) \xrightarrow{i_*} \pi_n(X, B) \xrightarrow{j_*} \pi_n(X; A, B) \xrightarrow{\partial} \pi_{n-1}(A, A \cap B) \xrightarrow{i_*} \ldots$$

The homomorphisms $i_*$ and $j_*$ are induced by the inclusion maps

$$i:(A, A \cap B) \rightarrow (X, B), \qquad j:(X; x_0, B) \rightarrow (X; A, B).$$

This sequence is one of the two homotopy sequences of the triad $(X; A, B)$; the other sequence is obtained by interchanging the roles of $A$ and $B$ throughout.

THEOREM 1. *The homotopy sequences of a triad are both exact.*

If $X = A \cup B$, then the inclusion map $i$ is an excision. It follows from exactness that $i_*$ is an isomorphism in all dimensions if and only if $\pi_n(A \cup B; A, B) = 0$ for all $n$. Thus if $\pi_n(A \cup B; A, B) \neq 0$ for some integer $n$, then the excision $i$ cannot induce isomorphisms onto in all dimensions, and the groups $\pi_n(A \cup B; A, B)$ are a "measure" of the amount of deviation from invariance under excision.

As an example of the application of these results, consider the triad $(S^n; E^n_+, E^n_-)$. It can be shown (see below) that $\pi_p(S^n; E^n_+, E^n_-) = 0$ for $p < 2n - 1$, and that $\pi_{2n-1}(S^n; E^n_+, E^n_-)$ is infinite cyclic. In this case the excision homomorphism $i_*: \pi_p(E^n_+, S^{n-1}) \rightarrow \pi_p(S^n, E^n_-)$ is equivalent to Freudenthal's "Einhängung";[2] more precisely, in the following diagram,

$$\begin{tikzcd}
 & \pi_{p-1}(S^{n-1}) \arrow[r, "E"] & \pi_p(S^n) \arrow[d, "k_*"] & \\
\cdots \arrow[r] & \pi_p(E^n_+, S^{n-1}) \arrow[u, "\partial"] \arrow[r, "i_*"] & \pi_p(S^n, E^n_-) \arrow[r] & \cdots
\end{tikzcd}$$

the boundary homomorphism $\partial$, and the homomorphism $k_*$ induced by an inclusion map, are isomorphisms onto, and the commutativity relation $i_* = k_* E \partial$ holds, where $E$ is the Einhängung. These facts, together with the exactness of the homotopy sequence of the triad, imply the Freudenthal theorems.

2. The homotopy groups of triads have many properties which are the analogs of familiar properties of the homotopy groups of spaces or pairs of spaces. We now list some of these properties:

(*a*) $\pi_n(X; A, B, x_0)$ is abelian for $n > 3$; simple examples show that it need not be so for $n = 3$.

(*b*) The system of groups $\pi_n(X; A, B, x)$ for $x \in A \cap B$ forms a local system of groups[3] in the space $A \cap B$ in the usual sense, and $\pi_1(A \cap B, x_0)$ is a group of operators on $\pi_n(X; A, B, x_0)$.

(*c*) If $A \supset B$, then $\pi_n(X; A, B)$ is isomorphic to $\pi_n(X, A)$, and the homotopy sequence of the triad $(X; A, B)$ reduces to that of the triple $(X, A, B)$.

(*d*) In case $\pi_3(X; A, B)$ is non-abelian, the kernels of the boundary homomorphisms $\partial: \pi_3(X; A, B) \rightarrow \pi_2(A, A \cap B)$ and $\partial': \pi_3(X; A, B) \rightarrow \pi_2(B, A \cap B)$ are both contained in the center of $\pi_3(X; A, B)$.

3. A topological space $X$ is said to be *n-connected* $(n > 0)$ if it is arcwise connected and $\pi_p(X) = 0$ for $1 \leq p \leq n$. A pair $(X, A)$ is said to be *n-connected* $(n \geq 1)$ if both $X$ and $A$ are arcwise connected, the natural homomorphism $\pi_1(A) \rightarrow \pi_1(X)$ is a homomorphism onto, and $\pi_p(X, A) = 0$, $2 \leq p \leq n$. Similarly, we will say a triad $(X; A, B)$ is *n-connected* $(n \geq 2)$ if both of the pairs $(A, A \cap B)$ and $(B, A \cap B)$ are 1-connected, and $\pi_p(X; A, B) = 0$, $2 \leq p \leq n$. By $\pi_2(X; A, B) = 0$ we mean that this set of homotopy classes of mappings contains a single element, the homotopy class of the constant map into $x_0$.

Let $(X^*, X)$ be a pair, and $\psi:(E^n, S^{n-1}) \rightarrow (X^*, X)$, $(n > 1)$, a continuous map which is a homeomorphism of $(E^n - S^{n-1})$ onto $X^* - X$. Let $\mathcal{E}^n = \psi(E^n)$, $\dot{\mathcal{E}}^n = \mathcal{E}^n \cap X = \psi(S^{n-1})$. We say that the space $X^*$ is obtained from $X$ by *adjunction of the cell* $\mathcal{E}^n$. Note that $(X^*; X, \mathcal{E}^n)$ is a triad, and $\dot{\mathcal{E}}^n$ is 0-connected.

THEOREM 2. *If the pair* $(X, \dot{\mathcal{E}}^n)$ *is m-connected,* $m \geq 1$, *then the triad* $(X^*; X, \mathcal{E}^n)$ *is* $(m + n - 1)$*-connected. (In case* $m = 1$, *it is also necessary to assume that* $\pi_2(X, \dot{\mathcal{E}}^n)$ *is abelian; in case* $n = 2$, *it is necessary to assume that X is simple relative to* $\dot{\mathcal{E}}^n$ *in all dimensions.)*

THEOREM 3. *If the space* $\dot{\mathcal{E}}^n$ *is m-connected,* $m \geq 1$, *then the boundary homomorphism*

$$\partial:\pi_{i+1}(X^*; X, \mathcal{E}^n) \rightarrow \pi_i(\mathcal{E}^n, \dot{\mathcal{E}}^n)$$

*is the zero homomorphism for* $2 \leq i \leq m + n - 1$.

The proofs of theorems 2 and 3 are a straightforward application of the theory of "obstructions"[4] to extensions and deformations of continuous mappings. By combining these theorems with the exactness of the homotopy sequence of the triad $(X^*; X, \mathcal{E}^n)$, important results are obtained as corollaries. Among these are the previously mentioned assertion that $\pi_p(S^n; E^n_+, E^n_-) = 0$ for $p < 2n - 1$, and an essential generalization of a theorem of J. H. C. Whitehead[5] about the homotopy groups of the pair $(X^*, X)$.

4. Let $f:(E^n; E^{n-1}_+, E^{n-1}_-) \rightarrow (X; A, B)$ be a map, and assume $u \in H^p(X, A)$, $v \in H^q(X, B)$, $p + q = n + 1$ and $u \smile v = 0$, where $u \smile v$ is the Alexander-Cech-Whitney cup product. In a recent paper,[6] Steenrod has shown how to associate with the elements $u$, $v$ an element $u \underset{f}{\smile} v$ of the factor group $H^n(E^n, S^{n-1})/f^*H^n(X, A \cup B)$. He has proved that the pairing thus defined, called the *functional cup product* is bilinear, and depends only on the homotopy class of $f$. Hence if $\alpha \in \pi_n(X; A, B, x_0)$ is the homotopy class of $f$, we can write $u \underset{f}{\smile} v = u \underset{\alpha}{\smile} v$ without ambiguity.

THEOREM 4. *If* $H^n(X, A \cup B) = H^{n+1}(X, A \cup B) = 0$, *and* $\gamma = \alpha + \beta$, $\alpha, \beta \in \pi_n(X; A, B, x_0)$, $n > 2$, *then*

$$u \underset{\alpha}{\smile} v + u \underset{\beta}{\smile} v = u \underset{\gamma}{\smile} v.$$

This conclusion can be alternately stated: *The functional cup product determines a homomorphism of the group* $\pi_n(X; A, B)$ *into the group of bilinear functions*

$$[H^p(X, A), H^q(X, B)] \rightarrow H^n(E^n, S^{n-1}).$$

In the important special case where $(X; A, B) = (S^n; E^n_+, E^n_-)$, and we use cohomology groups with integer coefficients, additional results can be obtained. It is easily seen that the group of bilinear maps

$$[H^n(S^n, E^n_+), H^n(S^n, E^n_-)] \to H^{2n-1}(E^{2n-1}, S^{2n-2})$$

is an infinite cyclic group. Denote this group by $G_n$. Then we can prove

THEOREM 5. *The homomorphism* $\pi_{2n-1}(S^n; E^n_+, E^n_-) \to G_n$ *determined by the functional cup product is an isomorphism onto.*

Now consider the following portion of the homotopy sequence of the triad $(S^n; E^n_+, E^n_-)$:

$$\pi_{2n-1}(S^n, E^n_-) \xrightarrow{j_*} \pi_{2n-1}(S^n; E^n_+, E^n_-) \xrightarrow{\partial} \pi_{2n-2}(E^n_+, S^{n-1}) \xrightarrow{i_*} \pi_{2n-2}(S^n, E^n_-).$$

Using the known properties of the functional cup product, additional information can be obtained about the homomorphisms $j_*$, $\partial$, and $i_*$.

(*a*) If $n$ is odd, then $j_* = 0$, $\partial$ is an isomorphism into, and image $\partial$ is an infinite cyclic subgroup of $\pi_{2n-2}(E^n_+, S^{n-1})$.

(*b*) If $n$ is even, and $\pi_{2n-1}(S^n)$ has an element of Hopf Invariant 1, then $j_*$ is onto, $\partial = 0$, and $i_*$ is an isomorphism onto. This case is known to occur for $n = 2, 4, 8$.

(*c*) If $n$ is even and $\pi_{2n-1}(S^n)$ has no element of Hopf Invariant 1, then it must have an element of Hopf Invariant 2. In this case $j_*$ is a homomorphism onto the subgroup of $\pi_{2n-1}(S^n; E^n_+, E^n_-)$ of index two, and image $\partial$ is a cyclic subgroup of $\pi_{2n-2}(E^n_+, S^{n-1})$, of order two. G. W. Whitehead[7] has recently announced that this case occurs when $n \equiv 2 \bmod 4$, $n > 2$.

These three statements are essentially the Freudenthal theorems, together with some improvements due to G. W. Whitehead.[7]

5. Let $A$ and $B$ be arcwise connected spaces, $a_0 \in A$, $b_0 \in B$, and let $A \vee B$ be the subspace of the cartesian product $A \times B$ defined by $A \vee B = (A \times b_0) \cup (a_0 \times B)$. $A \vee B$ is the union of $A$ and $B$ with a single point in common. Define $\mu_1 : A \to A \vee B$ by $\mu_1(x) = (x, b_0)$ and $\mu_2 : B \to A \vee B$ by $\mu_2(y) = (a_0, y)$, for $x \in A$, $y \in B$. Consider the following homomorphisms ($n > 1$):

$$\mu_{1_*} : \pi_n(A) \to \pi_n(A \vee B),$$
$$\mu_{2_*} : \pi_n(B) \to \pi_n(A \vee B),$$
$$\partial : \pi_{n+1}(A \times B, A \vee B) \to \pi_n(A \vee B).$$

We can paraphrase a result of G. W. Whitehead[7] as follows: The homomorphisms $\mu_{1_*}$, $\mu_{2_*}$, and $\partial$ are isomorphisms into, and $\pi_n(A \vee B)$ splits up into the direct sum

$$\mu_{1_*}\pi_n(A) + \mu_{2_*}\pi_n(B) + \partial\pi_{n+1}(A \times B, A \vee B).$$

From this it follows by an easy argument using exact sequences, that for $n > 2$, $\pi_n(A \vee B; A, B)$ is isomorphic to $\pi_{n+1}(A \times B, A \vee B)$, and that in the homotopy sequence of the triad $(A \vee B; A, B)$, the boundary operators are both trivial in all dimensions. If $A$ and $B$ are spheres, further results can be obtained. If theorems 2 and 3 are applied with $X^* = S^p \times S^q$, and $\mathcal{E}^n$ a $(p + q)$-simplex in a triangulation of $S^p \times S^q$, we find that $\pi_i(S^p \times S^q, S^p \vee S^q) \approx \pi_i(E^{p+q}, S^{p+q-1})$ for $i \leq p + q + \min(p, q) - 2$. It follows that $\pi_j(S^p \vee S^q; S^p, S^q) \approx \pi_j(S^{p+q-1})$ for $j \leq p + q + \min(p, q) - 3$.

Let $\varphi_n: (S^n; E^n_+, E^n_-) \to (S^n_1 \vee S^n_2; S^n_1, S^n_2)$ be the map defined by identifying all points of the "equator" $S^{n-1} \subset S^n$ to a single point. Then $\varphi_n$ induces a homomorphism of the corresponding triad homotopy sequences, as shown in the following diagram:

$$\begin{tikzcd}
\cdots \arrow[r, "\partial"] & \pi_r(E^n_+, S^{n-1}) \arrow[r, "i_*"] \arrow[d] & \pi_r(S^n, E^n_-) \arrow[r, "j_*"] \arrow[d] & \pi_r(S^n; E^n_+, E^n_-) \arrow[r, "\partial"] \arrow[d, "\varphi_{n_*}"] & \cdots \\
\cdots \arrow[r, "\partial"] & \pi_r(S^n_1) \arrow[r] & \pi_r(S^n_1 \vee S^r_2, S^n_2) \arrow[r] & \pi_r(S^n_1 \vee S^n_2; S^n_1, S^n_2) \arrow[r, "\partial"] & \cdots
\end{tikzcd}$$

Now $\pi_r(S^n) \approx \pi_r(S^n, E^n_-)$, and if $r \leq 3n - 3$, $\pi_r(S^{2n-1}) \approx \pi_r(S^n_1 \vee S^n_2; S^n_1, S^n_2)$ by the preceding paragraph. The composition of these two isomorphisms with the homomorphism $\varphi_{n_*} j_*$ of the diagram yields a homomorphism

$$H: \pi_r(S^n) \to \pi_r(S^{2n-1}), \ (r \leq 3n - 3).$$

It may be shown that this homomorphism is the same as the Generalized Hopf Homomorphism defined by G. W. Whitehead.[7] However, this definition is simpler and easier to work with than the original definition. Furthermore, we have demonstrated the existence of $H$ for $r \leq 3n - 3$, instead of for $r \leq 3n - 4$, as was originally done by Whitehead. By using Whitehead's methods, we can show that in the limiting case for $n = 3$, the homomorphism

$$H: \pi_6(S^3) \to \pi_6(S^5)$$

is a homomorphism onto. Since $\pi_6(S^5)$ is known to be cyclic of order two, this shows that $\pi_6(S^3) \neq 0$. An essential map of $S^6$ onto $S^3$ may be constructed as follows: Let $f: S^3 \times S^2 \to S^2$ be a map of type[8] $(\alpha, \iota)$, where $\alpha$ is a generator of $\pi_3(S^2)$ and $\iota \in \pi_2(S^2)$ is the class of the identity map. Now apply the Hopf construction[9] to $f$ to obtain a map $F: S^6 \to S^3$.

---

[1] Eilenberg, S., and Steenrod, N. E., these PROCEEDINGS, **31**, 117–120 (1945). Axiom 6 of this note is called the "excision axiom."

[2] Freudenthal, H., *Compositio Math.*, **5**, 299–314 (1937).

[3] Steenrod, N. E., *Ann. Math.*, **44**, 610–627 (1943).

[4] See Eilenberg, S., *Ann. Math.*, **41**, 231–251 (1940), and Blakers, A. L., *Bull. Am. Math. Soc.* **54**, abstract no. 413 (1948). Also a forthcoming paper by Olum, P., in *Ann. Math.* entitled "Obstructions to Extensions and Homotopies."

[5] Whitehead, J. H. C., *Proc. Lond. Math. Soc.*, **48**, 265, theorem 8 (1944).

[6] Steenrod, N. E., these PROCEEDINGS, **33**, 124–128 (1947). A complete exposition will appear soon in the *Ann. Math.*

[7] Whitehead, G. W., these PROCEEDINGS, 32, 188–190 (1946); also *Bull. Am. Math. Soc.*, 54, abstract no. 546 (1948). A complete exposition of these results will appear in *Ann. Math.* The authors are indebted to Professor Whitehead for allowing them access to his manuscript before publication.

[8] Whitehead, G. W., *Ann. Math.*, 47, 460–475 (1946).

[9] Hopf, H., *Fund. Math.*, 25, 427–440 (1935); also, Whitehead, G. W., *Ann. Math.*, 43, 636 (1942).

---

# ON THE "CENTRAL" PROBABILITY PROBLEM*

BY MICHEL LOÈVE

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA AT BERKELEY

Communicated by G. C. Evans, April 4, 1949

Let $\mathcal{L}(X)$, $\mathcal{L}(Y)$ be the probability laws of the real random variables (r. v.) $X$, $Y$, laws characterized either by the distribution functions (d. f.) $F(x)$, $G(x)$ or by their Fourier-Stieljes transforms—characteristic functions (c. f.) $f(u)$, $g(u)$. When $X$ or $Y$ possess subscripts, their d. f. and c. f. will have the same subscripts.

*I. The Problem.*—The problem of convergence towards a normal law of $\mathcal{L}\left(\sum_{k=1}^{n} X_k/c_n\right)$ when $n \to \infty$, $X_k$'s being independent r. v. and $c_n \to \infty$ being "sur" numbers, has played a central rôle in probability theory. It has been finally solved in 1934–1935 by Feller[4] and by P. Lévy.[8]

In its present formulation the "central" problem is that of limit laws, when $n \to \infty$, of sums

($\mathcal{P}$) $X_n = X_{n,1} + X_{n,2} + \ldots, + X_{n,\nu_n}$, $\nu_n \to \infty$, under the assumption of *independence*, i.e.

($\mathcal{I}$) for every fixed $n$, $X_{n,1}, X_{n,2}, \ldots, X_{n,\nu_n}$ are independent, and of *asymptotically and uniformly negligible components:*

($\mathcal{H}$) $\max\limits_{1 \leq k \leq \nu_n} P(|X_{n,k}| > \epsilon) \to 0$ for every given $\epsilon > 0$.

It has been shown that the class of limit laws coincides with that of infinitely divisible (i. d.) laws (Khintchine,[7] also Lévy,[8] after partial results of Bawly[1]) characterized, roughly speaking, as convolutions of a normal law and a continuum of Poisson laws (P. Lévy[8] after partial results of Kolmogoroff).[6] Conditions for convergence towards any i. d. law were given by Gnedenko[5] and Doeblin.[3]

The assumption ($\mathcal{H}$) is natural if one wants the limit laws to depend primarily upon the fact that the number $\nu_n$ of components of sums $X_n$ would be infinitely increasing. But the assumption ($\mathcal{I}$) is very restrictive. When ($\mathcal{I}$) is not assumed, the fundamental result for the primitive form of the problem is due to S. Bernstein:[2] if $E|X_k|^3 < \infty$, normal convergence holds under conditions which, in the case of independence, reduce to those of Liapounoff; for the general limiting process $(\mathcal{O})$, without any assumption about the moments, conditions which under $(\mathcal{S})$ reduce to those of Feller were given by the author.[9a]

The general problem of limit laws of sums in the case of dependence has not been considered until now and this work is an attempt in this direction. It appeared useful to extend the whole "central" problem to that of "asymptotic equivalence" (as. eq.) of laws, to be defined. This, in turn, led to the introduction of r. v. "in a wide sense," i.e., with possible positive probabilities for infinite values. In general, laws of two sums, varying with $n$, are compared for $n \gg 1$; if one of them reduces to a fixed law or converges towards it, the problem reduces to that of limit laws. In the case of one variable law $\mathcal{L}(X_n)$ it becomes that of the asymptotic behavior of $\mathcal{L}(X_n)$: its asymptotic type, if any, and more generally types of its accumulation laws, classes of as. eq. laws, etc.... The tools used, and some applications directly related to the problem of limit laws, which contain results obtained under $(\mathcal{S})$, are stated here.

It should be noticed that S. Bernstein's and the author's previous conditions for normal convergence (see also[9b]) appear as unnecessarily restrictive.

*II. The Tools.*—Let $H$-functions: $H(x)$, $H_n(x)$, ... be of uniformly bounded variation over $(-\infty, +\infty)$ and denote their Fourier-Stieljes transforms by $h(u)$, $h_n(u)$, .... Because of the meaning to be attached only to $\Delta H(x)$ and not to $H(x)$ itself, and the fundamental rôle of Fourier-Stieljes transforms, $H$-functions which, up to a constant, coincide at all their continuity points will be considered as equivalent and represented by the normalized elements of their equivalence classes:

$$H(-\infty) = \lim_{x \to -\infty} H(x) = 0, \qquad H(x) = \frac{H(x-0) + H(x+0)}{2}.$$

Any convergence of $H$-functions will be defined up to an equivalence. If not otherwise stated, limits will be considered for $n \to \infty$ and for all (finite) values of the arguments.

When $H_n(x)$ converges towards a limit function, this is an $H$-function: write $H_n(x) \to H(x)$—"simple" convergence. If, moreover, $\lim_{n \to \infty} H_n(+\infty) = H(+\infty)$, i.e., $\lim_{n \to \infty} \lim_{x + \infty} H_n(x) = \lim_{x \to +\infty} \lim_{n \to \infty} H_n(x)$, write $H_n(x) \xrightarrow{c} H(x)$—"complete" convergence (not to be confused with the one where var. $H_n \to$ var. $H$). The tools used are the following four lemmas of which $A$ and $B$ are extensions of the continuity theorem for c. f., $C$ is immediate and the basic lemma $D$ is elementary.

LEMMA A. *If $h_n(u)$ converges towards $k(u)$ almost everywhere then $H_n(x) \to$ some $H(x)$ and $h_n(u) = k(u)$ almost everywhere.*

**Lemma B.** *If $h_n(u)$ converges towards $k(u)$ continuous at the origin then $H_n(x) \xrightarrow{c} H(x)$.*

Let $S$ be a bounded set on the real line and $S'$ its complement with respect to $(-\infty, +\infty)$.

**Lemma C.**

$$|h(u)| \leqq |H(+\infty)| + 2\int_{S'} |dH(x)| + \sum_{j=1}^{m} \frac{|u|^j}{j!} \left| \int_S x^j dH(x) \right| + \frac{|u|^{m+1}}{j^{m+1}} \int_S \left| x^{m+1} dH(x) \right|.$$

Let $X_n = X_{n,1} + \ldots + X_{n,\nu_n}$, $Y_n = Y_{n,1} + \ldots + Y_{n,\nu_n}$, $Z_{n,k} = X_{n,1} + \ldots + X_{n,k-1} + Y_{n,k+1} + \ldots + Y_{n,\nu_n}$ and the prime denote conditional expressions given $Z_{n,k}$.

**Lemma D.** $|f_n(u) - g_n(u) \leqq \sum_{k=1}^{\nu_n} E|f'_{n,k}(u) - g'_{n,k}(u)|$.

*III. Asymptotic Equivalence.*—If $H_n(x) = F_n(x) - G_n(x) \xrightarrow{c} 0$ we say that $\mathfrak{L}(X_n)$ and $\mathfrak{L}(Y_n)$ are (completely) as. eq. and write $\mathfrak{L}(X_n) \overset{c}{\sim} \mathfrak{L}(Y_n)$. Put

$$a'_{n,k}(\epsilon) = \int_{|x|<\epsilon} x dF'_{n,k}(x),\ \sigma'^2_{n,k}(\epsilon) = \int_{|x|<\epsilon} [x - a'_{n,k}(\epsilon)]^2 dF'_{n,k}(x)$$

and $b'_{n,k}(\epsilon)$, $\tau'^2_{n,k}(\epsilon)$ when $F'_{n,k}(x)$ is replaced by $G'_{n,k}(x)$. Suppress the primes when they disappear in $F'_{n,k}(x)$, respectively, in $G'_{n,k}(x)$.

We state two results directly useful for the problem of limit laws and, to simplify, consider only usual—"strict" r. v.

I. *If, for $\epsilon \to 0$*

$$\overline{\lim_{n\to\infty}} \sum_k E \int_{|x|\geqq\epsilon} |dF'_{n,k}(x) - dG'_{n,k}(x)| = o(1),$$

$$\overline{\lim_{n\to\infty}} \sum_k E|a'_{n,k}(\epsilon) - b'_{n,k}(\epsilon)| = o(1),\ \overline{\lim_{n\to\infty}} \sum_k E|\sigma'^2_{n,k}(\epsilon) - \tau'^2_{n,k}(\epsilon)| = o(1),\ \overline{\lim_{n\to\infty}} \sum_k E\tau'^2_{n,k}(\epsilon) = O(\epsilon^{-1})$$

*then* $\mathfrak{L}(X_n) \overset{c}{\sim} \mathfrak{L}(Y_n)$.

Take $\mathfrak{L}'(Y_{n,k})$ to be the "conditional i. d. law" defined by

$$\log g'_{n,k}(u) = ia'_{n,k}(\epsilon)u + \int_{-\infty}^{+\infty} (\exp.\{iu[x - a'_{n,k}(\epsilon)]\} - 1)dF'_{n,k}(x)$$

II. *If, for $\epsilon \to 0$,*

$$\overline{\lim_{n\to\infty}} \sum_k E\left| \int_{-\infty}^{+\infty} (\exp.\{iu[x - a'_{n,k}(\epsilon)]\} - 1)dF'_{n,k}(x) \right|^2 = o(1)$$

*then*

$$\overline{\lim_{n \to \infty}} \ |f_n(u) - g_n(u)| = o(1).$$

*Under* ($\mathcal{H}$), *if*

$$\sum_k \int_{|x| \geqq \epsilon} dF_{n,k}(x) \leqq C_1 < \infty, \qquad \sum_k E\sigma'^2_{n,k}(\epsilon) \leqq C_2 < \infty$$

*then* $\mathfrak{L}(X_n) \overset{c}{\sim} \mathfrak{L}(Y_n)$.

Applied to the problem of limit laws these results give: *Asymptotic independence. If the hypotheses of I are satisfied with* $G'_{n,k}(x) = F_{n,k}(x)$ *then, under* ($\mathcal{H}$), *the necessary and sufficient conditions for* $\mathfrak{L}(X_n)$ *to converge towards a law, necessarily i. d. are, that for any* $\epsilon > 0$ $\sum\limits_k a_{n,k}(\epsilon) \to a(\epsilon)$

$\lim\limits_{\epsilon \to 0} \sum\limits_k \sigma^2_{n,k}(\epsilon) = \sigma^2$ *and*

$$\sum_k \int_{-\infty}^{-\epsilon} dF_{n,k}(x) \to \int_{-\infty}^{-\epsilon} dL(x), \ \sum_k^{\infty} \int_{\epsilon}^{\infty} dF_{n,k}(x) \to \int_{\epsilon}^{\infty} dL(x)$$

*where* $a(\epsilon)$, $\sigma^2$ *and* $L(x)$ *characterize the i. d. limit law.* In the case of independence this becomes the Gnedenko-Doeblin result.

Let us give two simple consequences; $\epsilon > 0$ is arbitrarily small.

*Examples: 1. If*

$$\sum_k \int_{|x| \geqq \epsilon} dF_{n,k}(x) \to 0,$$

$$\sum_k E|a'_{n,k}(\epsilon) - a_{n,k}(\epsilon)| \to 0, \ \sum_k a_{n,k}(\epsilon) \to a$$

$$\sum_k E|\sigma'^2_{n,k}(\epsilon) - \sigma^2_{n,k}(\epsilon)| \to 0, \qquad \sum_k \sigma^2_{n,k}(\epsilon) \to \sigma^2$$

*then* $\mathfrak{L}(X_n)$ *converges towards the normal law* $N(a, \sigma^2)$—contains the S. Bernstein's and author's conditions.

Let $S = (-\epsilon, +\epsilon) \cup (1 - \epsilon, 1 + \epsilon)$.

*2. If* $\sum\limits_k E \int\limits_{|x-1|<\epsilon} |dF'_{n,k}(x) - dF_{n,k}(x)| \to 0, \sum\limits_k \int\limits_{|x-1|<\epsilon} dF_{n,k}(x) \to \lambda$, $\sum\limits_k E|a'_{n,k}(\epsilon)| \to 0, \sum\limits_k E\sigma'^2_{n,k}(\epsilon) \to 0, \sum\limits_k \int\limits_{S'} dF_{n,k}(x) \to 0$, *then* $\mathfrak{L}(X_n)$ *converges towards the Poisson law with parameter* $\lambda$—in the case of independence these conditions reduce to those of Gnedenko-Marcinkiewicz.

*IV. Extensions.*—1. Let now $\nu_n$ be a r. v. Put $P(\nu_n = r) = p_n(r)$, $P(\nu_n \leqq r) = P_n(r)$, denote by the subscript $r$ the expectations given $\nu_n = r$, assume that the quantities below exist (they certainly do when $E\nu_n < \infty$), and write $\mathcal{E}_{n,k}$ for $\sum\limits_{r \geqq k} p_n(r) E_r$. Notice that to the condition $\nu_n \to \infty$ when $\nu_n$ is "sur" would correspond, for instance, the condition: $\mathfrak{L}(Y_n)$ converges towards a law (in a wide sense) degenerated at infinity.

*The results above hold for $\nu_n$ random if we replace $E$ by* $\mathcal{E}_{n,k}$ (observe that $F_{n,k}(x) = EF'_{n,k}(x)$).

*If the $X_{n,k}$'s do not depend upon $\nu_n$, then we have only to replace them by the r. v. in a wide sense with conditional laws $P_n(k)F'_{n,k}(x)$ instead of $F'_{n,k}(x)$.*

2. *The results above hold for random vectors $X_{n,k}$, $Y_{n,k}'$ ... with the corresponding interpretation of the symbols used.*

It is hoped that the proofs and various details will be published in the Statistical Series of the University of California.

---

* This work has been supported in part by the Office of Naval Research.

[1] Bawly, G., *Matem. Sbornik.*, 1 (43), 917–929 (1936).

[2] Bernstein, S., *Math. Ann.*, **97**, 1–59 (1927).

[3] Doeblin, W., *Bull. Sc. Math.*, **63**, 22–32, 35–64 (1939).

[4] Feller, W., *Math. Zeitschr.*, **40**, 521–539 (1935); **42**, 301–312 (1937).

[5] Gnedenko, B., *DOKL.* **18**, 231–234 (1938); **22**, 886–871 (1939). *IZV. A. N.* 181–232, 643–647 (1939).

[6] Kolmogoroff, A., *Rend. Lincei*, **15**, 805–808, 866–869 (1932).

[7] Khintchine, A., *Matem. Sbornik*, **2** (44), 79–119 (1937).

[8] Lévy, P., *Théorie de l'Addition des Variables Aléatoires*, Paris, Gauthier-Villars (1937).

[9a] Loève, M., *J. de Math.*, **24**, 249–318 (1945).

[9b] Loève, M., *C. R. Ac. Sc. Paris*, **227**, 1335–1337 (1948).

[10] Marcinkiewicz, J., *Fund. Math.*, **30**, 354 (1938).

---

# ON CONGRUENCES WHICH RELATE THE FERMAT AND WILSON QUOTIENTS TO THE BERNOULLI NUMBERS

By H. S. Vandiver

Department of Applied Mathematics, University of Texas

Communicated April 30, 1949

In a recent paper[1] the writer proved a result which may be written in the following form

$$-q\left(\frac{m}{r}\right) + W - \sum_{n=1}^{p-2} \left(\frac{m}{r}\right)^n \frac{b_n}{n} \equiv \frac{\lambda(m, r)}{r}. \qquad (1)$$

modulo $p$, where $m > 0$, $0 < r < p$, $(m, p) = 1$, $(k, p) = 1$, $p$ is an odd prime,

$$q(s) = \frac{s^{p-1} - 1}{p},$$

$$W = \frac{(p-1)! + 1}{p},$$

$$\lambda(m, r) = \sum{}' \frac{\rho^r - 1}{\rho^k - 1}; \; m > 1,$$

with $\lambda(1, r) = 0$, $\rho$ an $m$th root of unity $\neq 1$, the sign $\sum'$ means summation over all $m$th roots of unity $\neq 1$, and $b_n$ is the Bernoulli number defined by the recursion formula $(b + 1)^k = b_k$ for every $k > 1$, where the left-hand member is expanded and $b_t$ substituted for $b^t$ in the result.

In the present paper we deduce some new relations concerning the Fermat and Wilson quotients and the Bernoulli numbers. The principal results are (6a), (9), (11), (11a) and (22).

To show that (1) is equivalent to the other relation referred to, we note that the $\beta(m, r)$ of our paper $C$ satisfies

$$\lambda(m, r) = \beta(m, r) - \sum{}' \frac{1}{\rho^p - 1} \tag{2}$$

and applying the known relation

$$\sum{}' \frac{1}{\rho^p - 1} = \sum{}' \frac{1}{\rho - 1} = \sum{}' \frac{\rho + 2\rho^2 + \ldots + (m-1)\rho^{m-1}}{m},$$

gives

$$\sum{}' \frac{1}{\rho^p - 1} = \frac{1 - m}{2} = \frac{1}{2} - \frac{m}{2},$$
$$= \frac{1}{2} - (-mb_1).$$

We also have

$$q\left(\frac{m}{r}\right) = q(m) - q(r),$$

by the definition of $q$. We now consider some of the properties of $\lambda(m, r)$. Obviously

$$\lambda(m, r) = \lambda(m, r + tm). \tag{3}$$

Set $ps \equiv r \pmod{m}$, $m > 1$, $0 < s < m$, then

$$\lambda(m, r) = \sum{}' \frac{\rho^r - 1}{\rho^p - 1} = \sum{}' \frac{\rho^{ps} - 1}{\rho^p - 1},$$
$$= \sum{}' \frac{\rho^s - 1}{\rho - 1},$$
$$= m - s;\ m > 1. \tag{4}$$

Also as $r$ ranges over the integers $1, 2, \ldots, m - 1$, then the corresponding numbers $s$ range over the same set in some order. Hence

$$\sum_{r=1}^{m-1} \lambda(m, r) = (m - 1)\left(m - \frac{m(m-1)}{2}\right),$$

$$= \frac{m(m-1)}{2}. \tag{5}$$

In particular if $p \equiv 1 \pmod{m}$, $\lambda(m, r) = m - r$. Also $\lambda(m, r) = 0$ if $r \equiv 0 \pmod{m}$.

We now deduce some relations from (1). Set $m = 1$ and $r = 1, 2, \ldots, p - 1$, in turn and add the resulting congruences, then since

$$\sum_{m=1}^{p-1} r^a \equiv 0 \pmod{p},\ a < p - 1, \tag{5a}$$

we have the known relation

$$W \equiv \sum_{r=1}^{p-1} q(r) \pmod{p}. \tag{6}$$

Now set in (1), $r = 1$ and add the relations found by setting $m = 1, 2, \ldots, p - 1$ in turn we have[2] using (5a) and (6)

$$-2W \equiv \sum_{m=1}^{p-1} \lambda(m, 1) \pmod{p}. \tag{6a}$$

Multiplying (1) through by $r$, the result may be written

$$-rq(m) + rq(r) + rW - \sum m^n r^{p-n} \frac{b_n}{n} \equiv \lambda(m, r) \pmod{p}. \tag{7}$$

Multiply this through by $m^k$, and add the results for $m = 1, 2, \ldots, p - 1$ we find from (5a), using the known relation (C, p. 109),

$$\sum_{a=1}^{p-1} a^t q(a) \equiv -\frac{b_t}{t} \pmod{p}, \tag{8}$$

the congruence

$$-r\frac{b_k}{k} - r^{k+1}\frac{b_{p-1-k}}{p-1-k} \equiv \sum_{m=1}^{p-1} m^k \lambda(m, r) \pmod{p}. \tag{9}$$

Now, taking $m = 1, 2, \ldots, p - 1$ in (7) as it stands, and adding, we have by (6) a result which may be written

$$q(r) + 2W \equiv -\sum_{m=1}^{p-1} \frac{\lambda(m, r)}{r} \pmod{p}, \tag{10}$$

and by using (6a) we have

$$q(r) \equiv -\sum_{m=1}^{p-1} \frac{\lambda(m, r)}{r} + \sum_{m=1}^{p-1} \lambda(m, 1) \pmod{p}. \tag{11}$$

On the surface, at least, this relation throws little if any light on the properties of $q(r)$, but on the other hand the known properties of $q(r)$ enable us to find relations involving $\lambda(m, r)$ alone for various values of $m$ and $r$. For example since, if $rt < p$,

$$q(rt) \equiv q(r) + q(t) (\text{mod } p),$$

we have from (11),

$$\sum_{m=1}^{p-1} \left(\frac{\lambda(m, r)}{r} + \frac{\lambda(m, t)}{t}\right) \equiv \sum_{m=1}^{p-1} \left(\frac{\lambda(m, rt)}{rt} + \lambda(m, 1)\right), \quad (11a)$$

modulo $p$.

A simpler property of the $\lambda(m, r)$ is given by

$$\lambda(m, m - r) = m - \lambda(m, r), \quad (12)$$

in view of (4).

Prof. Carlitz, after looking over our paper C, pointed out to me that the formula (6a) of that paper could be derived immediately from the known formula

$$(b - 1)^n - b_n = n(-1)^n, \quad (13)$$

by putting $n = p - 1$, and using

$$\binom{p-1}{i} \equiv (-1)^i (\text{mod } p). \quad (13a)$$

This proof is even simpler than the one referred to in the second footnote of C. He also showed that from (13) of the present paper, after setting $n = p$, we may obtain a result which is the formula (20) of C for $m = 1$, $r = 1$. By an extension of his method it is possible to prove a generalization of this formula (20) as we shall later show. In the original proof of (20) of C a number of known theorems concerning Wilson's quotient, Fermat's quotient and the Bernoulli numbers were employed, whereas in the present proof we shall see, aside from Fermat's theorem, that the only known results employed are (21a) together with

$$(b + t)^n - b_n = n(1^{n-1} + 2^{n-1} + \ldots + (t - 1)^{n-1} \quad (14)$$

and[4]

$$\frac{pb_{p-1} + 1}{p} \equiv W + 1 \, (\text{mod } p). \quad (15)$$

After (15) of C or (1) of the present paper is thereby proved, a number of the theorems employed in the first proof will follow as corollaries. The argument depends on finding two expressions for

$$\frac{(mb + k)^p}{p},$$

modulo $p$, where, as in the definition of $b_1, b_2, \ldots b_n$, we expand $(mb + k)^p$ by the binomial theorem and set $b_h$ for $b^h$ in the result. We have, if $k$ can be expressed in the form $mt + ps$, with $s$ and $t$ integers,

$$\begin{aligned}(mb + k)^p &= ((mb + mt) + ps)^p \\ &= (mb + mt)^p + p^2s(mb + mt)^{p-1} + \binom{p}{2} p^2s^2 \\ &\qquad (mb + mt)^{p-2} + \ldots. \end{aligned} \tag{16}$$

Now, in this development, every term after the second is divisible by $p^2$ since $p$ does not occur in the denominator of any Bernoulli number found in the expansion of $(mb + mt)^n$ for $n < p - 1$. This follows from the definition of the $b$'s. Hence we may write

$$(mb + k)^p \equiv (mb + mt)^p + p^2s(mb + mt)^{p-1} \tag{17}$$

modulo $p^2$. Now in the same way

$$p^2s(mb + mt)^{p-1} \equiv p^2sm^{p-1}b_{p-1} \equiv pb_{p-1}\cdot ps \equiv -ps,$$

modulo $p^2$, using $pb_{p-1} \equiv -1 \pmod{p}$, which follows from (15). Also

$$\begin{aligned}(mb + mt)^p &= m^p(b + t)^p, \\ &= m^pb^p + pm^p(1^{p-1} + 2^{p-1} + \ldots + (t - 1)^{p-1}), \\ &\equiv pm\left(t - 1 - \left[\frac{t}{p}\right]\right), \end{aligned} \tag{18}$$

modulo $p^2$, for $t > 0$, using $b_p = 0$ for $p$ odd. Hence (17) gives, $[x]$ being the greatest integer in $x$, $(mb + k)^p \equiv -ps + pmt - pm - pm\left[\frac{t}{p}\right]$, modulo $p^2$, and since $mt = k - ps$

$$\frac{(mb + k)^p}{p} \equiv -s - m + k - m\left[\frac{t}{p}\right] \pmod{p}. \tag{19}$$

On the other hand we may write by ordinary expansion

$$(mb + k)^p = p(mb)^{p-1} + k^p + pM, \tag{20}$$

where

$$M = \sum_{n=1}^{p-2} \binom{p}{n} m^n b_n k^{p-n}.$$

Now

$$\frac{pm^{p-1}b_{p-1}k + k^p}{p} = \frac{k(pb_{p-1}(m^{p-1} - 1) + pb_{p-1} + k^{p-1} - 1)}{p},$$

$$\equiv k\left(pb_{p-1}q(m) + q(k) + \frac{pb_{p-1} + 1}{p}\right),$$

$$\equiv k(-q(m) + W + 1 + q(k)), \qquad (21)$$

modulo $p$, by (15) and noting that from (15) we obtain $pb_{p-1} \equiv -1 \pmod{p}$, with $(m, p) = 1$, $(k, p) = 1$.

We selected, in the above argument, $k$ such that $k = mt + ps$ with $t > 0$. We may select $t > 0$ such that $mt \equiv k \pmod{p}$, since $(m, p) = 1$. Hence $k$ may be any integer such that $(k, p) = 1$. Hence from (19), (20) and (21) we have, noting that

$$\binom{p}{n} \equiv (-1)^{n+1}\frac{p}{n} \pmod{p^2};\ q\left(\frac{m}{k}\right) - W + \sum_{n=1}^{p-2}\left(\frac{m}{k}\right)^n \frac{b_n}{n} \equiv \frac{1}{k}\left(s + \left[\frac{t}{p}\right]\right). \qquad (22)$$

modulo $p$, where $(m, p) = 1$, $(k, p) = 1$, $k = ps + mt$, $s$, $m$ and $t$ are integers, $t > 0$, $q(x)$, $W$ and $b$ defined as in connection with (1).

The relation (22) is an extension of (1). For if we introduce the conditions stated in connection with (1), namely, $m > 0$, $0 < r < p$, and set $k = r$ in (22) then in $r = ps + mt$, $s < 0$ since $t > 0$, so if $s_1$ is now the original $s$ in (4) we have, since $0 < s_1 < m$; $s_1 = m + s$ which reduces (22) to (1). By the use of (22) such relations as (11a) may be generalized.

We close with a remark of an elementary nature. In the relation (22) we reduced binomial coefficients modulo $p$. In a similar way we obtain, if

$$\binom{0}{0} = 1, \binom{a}{b} = 0 \text{ for } b > a,$$

$$(-1)^a\binom{p-1-a}{b} \equiv (-1)^b\binom{p-1-b}{a},$$

modulo $p$, for $0 \leq b < p$; $0 \leq a < p$. This is a "law of reciprocity" which I have not observed in the literature.

---

[1] These PROCEEDINGS, 34, 103–110 (1948). This paper will be referred to as C.

[2] For other properties of Wilson's quotient cf. Emma Lehmer, *Ann. Math.*, 39, 352 (1938).

[3] One proof of (15) is given in Vandiver, *Duke Math. Jour.*, 8, 578 (1941), as a special case of a more general theorem. Other proofs are known.

# SECOND ORDER DIFFERENTIAL EQUATIONS OF RANK 2 IN PHYSICS

BY EDWARD KASNER AND DON MITTLEMAN

DEPARTMENT OF MATHEMATICS, COLUMBIA UNIVERSITY

Communicated March 30, 1949

*1. Introduction.*—All one-parameter families of plane curves are equivalent under point transformation. Thus, under the group of point transformations in the plane, there is no invariant of a one-parameter family of curves. However, under this same group, two-parameter families of curves (differential equations of second order) do possess invariants. Kasner[1] has determined the fundamental absolute invariants of differential elements of second order at a point. In the course of this investigation it was found expedient to introduce the homogeneous coördinates

$$z_1:z_2:z_3:z_4:z_5 = 1:y':y'^2:y'^3:y''.$$

This suggested the classification of differential equations of the second order, algebraic in $y'$ and $y''$, according to the degree of the corresponding equation in the $z$ coördinates. To an equation of first degree

$$A_1z_1 + A_2z_2 + A_3z_3 + A_4z_4 + A_5z_5 = 0$$

there corresponds a differential equation of the form

$$Ey'' = Ay'^3 + By'^2 + Cy' + D,$$

where the coefficients are functions of $x$, $y$. This equation of rank 1 has been extensively studied by Lie, R. Liouville, Tresse and Kasner. It includes, among others, the differential equations for the geodesics of a surface, velocity families, natural families, isogonal families and $\Gamma$ families.

The *rank* of a differential equation of second order, algebraic in $y'$ and $y''$, is defined as the degree of the corresponding $z$ equation.

*The rank of a differential equation is invariant under an arbitrary point transformation.*

The differential equations of rank 2 are:

$$A_0y''^2 + (B_0 + B_1y' + B_2y'^2 + B_3y'^3)y'' + (C_0 + C_1y' + C_2y'^2 + C_3y'^3 + C_4y'^4 + C_5y'^5 + C_6y'^6) = 0,$$

where the coefficients are arbitrary functions of $x$, $y$.

*2. Central Loci.*—Through a fixed point there pass integral curves of a given differential equation. The locus of the centers of the circles of curvature, constructed at a fixed point, to the integral curves passing through the point, is termed the *central locus*. With the fixed point as origin, the central locus, for equations of rank 1, is the cubic

$$E(x_0^2 + y_0^2) = -Ax_0^3 + Bx_0^2y_0 - Cx_0y_0^2 + Dy_0^3.$$

Such a cubic is characterized by the fact that the origin (the fixed point) is a double point and the two tangents are the minimal lines.

The only non-trivial degeneracy of this cubic occurs when $A = C$ and $B = D$, in which case the central locus is a straight line. A two-parameter family of curves for which, at each point, the central locus is a straight line is a *velocity family* (*of rank 1*).

For equations of rank 2, the central locus of the $2\infty^1$ integral curves through a fixed point is the special sextic:

$$A_0[x_0^2 + y_0^2]^2 + (B_0y_0^3 - B_1x_0y_0^2 + B_2x_0^2y_0 - B_3x_0^3)[x_0^2 + y_0^2] + (C_0y_0^6 - C_1x_0y_0^5 + C_2x_0^2y_0^4 - C_3x_0^3y_0^3 + C_4x_0^4y_0^2 - C_5x_0^5y_0 + C_6x_0^6) = 0.$$

This sextic is characterized by the fact that the origin (the fixed point) is a quadruple point, the tangent lines are the two minimal lines, each counted twice, and they intersect the sextic only at the origin.

There is non-trivial degeneracy of the sextic when $C_3 = C_1 + C_5$ and $C_0 + C_4 = C_2 + C_6$. In this case the central locus is the special quartic characterized by the fact that the origin is a double point and the two tangents are the minimal lines. Further degeneracy occurs if $B_0 = B_2$, $B_1 = B_3$, $C_1 = C_5$, and $C_0 + 2C_6 = C_4$. In this case, the central locus is the conic:

$$A_0 + B_0y_0 - B_1x_0 + C_6x_0^2 - C_5x_0y_0 + C_0y_0^2 = 0. \qquad (1)$$

An important subset of differential equations of rank 2 for which the central loci are conics is that for which, at each point, the focus of the conic is at the point. In this case, the central loci are *focal conics* and the integral curves are termed a *velocity family of rank 2*. The conditions that the focus of the conic (1) be at the origin are $B_0B_1 = 2A_0C_5$ and $B_0^2 - B_1^2 = 4A_0(C_0 - C_6)$.

The differential equations of the $2\infty^2$ curves of a velocity family of rank 2 are

$$y'' = (\mathfrak{A} + \mathfrak{B}y')(1 + y'^2) \pm \mathfrak{C}(1 + y'^2)^{3/2},$$

where $\mathfrak{A}$, $\mathfrak{B}$ and $\mathfrak{C}$ are functions of $x$, $y$.

*3.* Differential equations of rank 2 arise from many diverse problems in geometry and physics. It is proposed to indicate some of these.

*4.* *Scale Curves.*[2]—Let T be a non-conformal mapping of the surface $\Sigma$, whose first fundamental form is

$$dS^2 = E(x, y)dx^2 + 2F(x, y)dx\,dy + G(x, y)dy^2,$$

upon the plane $\pi$, for which $ds^2 = dx^2 + dy^2$. The curves, on $\Sigma$ or $\pi$, along which $\sigma^2 = (dS/ds)^2$ is constant, are scale curves. The totality of these scale curves, for a specific mapping **T**, satisfies a differential equation of rank 2.

*5.* The two-parameter family of curves whose finite equation is either of the two distinct types

$$f_1a^2 + f_2a + f_3b + f_4 = 0$$
$$f_1ab + f_2a + f_3b + f_4 = 0,$$

where the $f$'s are functions of $x$, $y$ and $a$, $b$ are the parameters, satisfies a differential equation of rank 2.

*6.* Any set of $\infty^2$ curves which is equivalent under a point transformation to the $\infty^2$ circles of unit radius satisfies a differential equation of rank 2.

Any set of $\infty^2$ curves which is conformally equivalent to the circles of unit radius is a velocity family of rank 2.

*7.* If $\nu(xy)$ is a scalar point function, **A** a vector point function, $d\mathbf{s}$ the vector arc length, $ds$ the scalar arc length, then the system of curves which are extremals of the variational problem

$$\delta\int(\nu ds + \mathbf{A}\cdot d\mathbf{s}) = 0$$

is a velocity family of rank 2.

*8.* *Electric Particle Trajectories.*[3]—An electrically charged particle moves in a static electromagnetic field. If the motion is planar, the totality of trajectories, corresponding to all possible initial conditions, constitutes a three-parameter family of curves. If the electrostatic field is derivable from a potential, these curves may be grouped into $\infty^1$ families of $\infty^2$ trajectories each. Each family of $\infty^2$ trajectories is a velocity family of rank 2.

*9.* *Motion in a Rotating Plane and the Restricted Problem of Three Bodies in the Plane.*[4]—Let there be defined in the plane two rectangular Cartesian coördinate systems which have a common origin but one of which rotates, with respect to the other, with constant angular velocity $\omega$. In the rotating coördinate system, the equations of motion for a particle of unit mass are

$$\ddot{x} - 2\omega\dot{y} = F \qquad (2)$$
$$\ddot{y} + 2\omega\dot{x} = G,$$

where $F$ and $G$ may be thought of as representing the components of a force in the $x$ and $y$ directions, respectively. In general, $F$ and $G$ depend on $x$, $y$, $\dot{x}$, $\dot{y}$ and $t$. The field will be called *positional* when $F$ and $G$ depend only on $x$, $y$. The totality of trajectories of a particle moving in such a positional field constitutes a three-parameter family of curves. If the field is derivable from a potential $-H(xy)$, then the equations (2) admit a first integral

$$\dot{x}^2 + \dot{y}^2 + 2H = 2E,$$

where $E$ is a constant of integration. In this case, the set of $\infty^3$ trajectories is composed of $\infty^1$ families of $\infty^2$ trajectories each, each family being associated with a specific value of $E$. The family of trajectories so determined is a velocity family of rank 2.

The classic example of the *restricted problem of three bodies* in the plane lies within this category.

*10.* The reason for calling a two-parameter family of curves whose central loci are focal conics a velocity family of rank 2 is made apparent by the following considerations.

In a plane, the intrinsic equations of motion of a particle of unit mass are

$$\frac{d^2s}{dt^2} = T \tag{3}$$

$$\frac{v^2}{r} = N, \tag{4}$$

where $s$ is the arc length, $r$ the radius of curvature of the trajectory, $v$ the velocity, $t$ the time, and $T$ and $N$ the tangential and normal components of the force. A trajectory is uniquely determined by the initial position, initial direction and initial speed. For the actual trajectory, the speed varies from point to point. If, in equation (4), the square of the speed, $v^2$, is replaced by a constant, say $c^{-1}$, the resulting equation is

$$\gamma = cN, \tag{5}$$

where, of course, $\gamma$ is curvature. The curves which satisfy this differential equation are not, in general, trajectories. When the normal component of the force, $N$, is of the form

$$N = \varphi \cos \mu + \psi \sin \mu,$$

where $\mu$ is the inclination of the normal to the curve, $\varphi$ and $\psi$ point functions, and the constant $c$ equals unity, the two-parameter family of curves so defined by (5) has been called by Kasner[5] a velocity family (of rank 1).

When $N$ is of the form

$$N = \varphi \cos \mu + \psi \sin \mu + \chi,$$

where $\chi$ is a point function and the other symbols retain their meaning, the two-parameter family of curves defined by (5) is a velocity family of rank 2.

[1] Kasner, E., "The Geometry of Differential Elements of the Second Order with Respect to the Group of All Point Transformations," *Am. J. Math.*, **28**, (3), 203–213 (1907).

[2] Kasner, E., and DeCicco, J., "Scale Curves in General Cartography," PROC. NAT. ACAD. SCI., **30**, 211–215 (1944). "General Theory of Scale Curves," *Am. J. Math.*, **68**, No. 1 (1946).

[3] Lorentz, H. A., *Enc. d. math. Wiss.*, **14**, 156. Pauli, W., Jr., *Ibid.*, **19**, 634.

[4] Moulton, F. R., *An Introduction to Celestial Mechanics*, Revised, The Macmillan Company, New York, 1914, pp. 277–281.

[5] Kasner, E., "Differential-Geometric Aspects of Dynamics," *Am. Math. Soc. Colloquium Publication*, Vol. 3 (1912). A series of papers in the *Trans. Am. Math. Soc.*, Vols. 7–11 (1906–1911). Kasner, E., and DeCicco, J., "The Geometry of Velocity Systems," *Bull. Am. Math. Soc.*, **49** (4) (1943).

PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES
Volume 35 July 15, 1949 Number 7

# *ANTIBIOTIC SUBSTANCES FROM BASIDIOMYCETES. IV. MARASMIUS CONIGENUS**

BY FREDERICK KAVANAGH, ANNETTE HERVEY AND WILLIAM J. ROBBINS

DEPARTMENT OF BOTANY, COLUMBIA UNIVERSITY, AND THE NEW YORK BOTANICAL GARDEN

Communicated May 10, 1949

In a survey of Basidiomycetes for antibacterial activity, *Marasmius conigenus* (No. 6890) was reported to inhibit the growth of *Staphylococcus aureus* when tested by the strip or the "hole" method.[1] Dr. W. H. Wilkins kindly supplied us with a subculture of this organism with which we continued studies. Other species of *Marasmius* produce antibacterial substances.[2, 3]

When grown on corn steep, thiamine peptone, or potato dextrose agar, and tested by the streak method,[4, 5] *Marasmius conigenus* showed in our hands marked antibacterial activity against *Staphylococcus aureus* (H) slight activity against *Escherichia coli* and none for *Mycobacterium smegma.*

Considerable activity against *Staph. aureus* was observed with discs[4, 5] cut from fungous cultures grown on the three media. The position of the disc in relation to the fungous colony determined the size and character of the inhibition zone. The disc cut from the oldest part of the colony produced the largest inhibition zone which was clear near the disc and surrounded by a zone of partial inhibition. Discs cut at a distance of 20 mm. from the edge of the fungous colony gave zones of partial inhibition. Isolations from these zones have not yielded strains of *Staph. aureus* resistant to the antibacterial material from this fungus.

*Antibacterial Activity of Culture Liquids.*—The fungus was grown at 25°C. in 2800-ml. Fernbach flasks containing beech shavings and a corn steep medium.[5] In about four weeks, when the surface of the liquid was about one-half covered with mycelium, the activity of the culture liquid against *Staph. aureus* was 256 or 512 dilution units per milliliter. Reflooding such mats with fresh medium produced liquids of as high activity one week or more after reflooding. The culture liquid from a seven-month old mat had an activity of 256 dilution units per milliliter 19 days after reflooding. Although year-old mats still produce culture liquids with an activity of 64 dilution units or more, the mats were usually not kept longer than five or six months as the increased thickness of mycelium made reflooding less practical. Mats were successfully reflooded six times.

The culture liquid in serial dilution evidenced little activity against *Escherichia coli* and the acid-fast bacteria as represented by *Mycobacterium phlei* and *M. smegma.*

The culture liquid lost its antibacterial activity at pH 8.5 and higher within an hour at 25°C. For the samples tested, there was no loss of activity at pH 3 on boiling for 10 minutes.

*Isolation of a Crystalline Antibacterial Substance.*—A white crystalline substance with antibacterial activity was isolated from the culture liquid of *Marasmius conigenus*. The procedure followed in the isolation of this compound may be illustrated by an example.

Ten liters of culture liquid with an activity of 64 dilution units per milliliter were acidified to pH 2 to 3 with hydrochloric acid and extracted twice, each time with one-tenth of its volume of chloroform. The chloroform was separated from the aqueous phase by passage through a de Laval separator. The chloroform was evaporated *in vacuo* to 25 ml. It contained 1373 mg. of dry matter. It was extracted twice, each time with 25 ml. of phosphate buffer at pH 5.8. The chloroform solution after extraction with buffer contained 1320 mg. of solids with an activity of 512 dilution units per milligram. The chloroform was removed by distillation *in vacuo;* the solids were dissolved in ether and extracted with 2% sodium bicarbonate solution. The bicarbonate removed 974 mg. of solids. The bicarbonate solution was acidified and extracted with ether. When the ether was evaporated slowly, white crystals mixed with a brown gum were formed. The gum was removed by washing carefully with ether. The crystals were purified by recrystallization from ether and additional crystals obtained from the gummy residue. The yield was 453 mg. of crystals which formed long white needles when crystallized from ether, dilute aqueous alcohol or acetone, water or ether upon the addition of hexane.

The maximum yield of crystals obtained was 100 mg. per liter of culture liquid. We have named this crystalline substance marasmic acid. Our observations indicate that culture liquids of 64 to 512 dilution units per milliliter all contained approximately the same amount of marasmic acid per liter (about 100 mg.).

*Chemical Properties of Marasmic Acid.*—Marasmic acid was very soluble in acetone, soluble in ethanol, ether, chloroform and water (1.1 mg./ml. at 25°) but insoluble in hexane.

The melting point in an open capillary tube was 165°C. (uncorr.). There was a temperature range through which part of the substance decomposed and the rest of the crystals dissolved in the melt. Determined in a capillary sealed off *in vacuo*, the melting point was sharp at 174–175°C. corr. without decomposition; it remelted at the same temperature even after the second or third melting.

The crystals titrated in 30% alcohol as a monobasic acid with a pK of 5.8 and an equivalent weight of 276. Qualitative tests did not indicate the presence of nitrogen, sulfur or halogens.

The quantitative analysis[6] was:

| | C | H | MOL. WT. |
|---|---|---|---|
| Found | 68.21 | 7.17 | |
| Computed for $C_{16}H_{20}O_4$ | 69.6 | 7.25 | 276 |

The optical activity of marasmic acid was $[\alpha]_D^{25} = +176$ (1.4% in acetone).

Marasmic acid in alcohol had absorption peaks in the ultra-violet region at from 240 to 242 mμ with a molecular extinction coefficient $\epsilon = 10{,}400$ and at 314–318 mμ with $\epsilon = 45.6$. In phosphate buffers of increasing pH, the position of the main peak was shifted slightly toward the red and the extinction decreased. In pH 1.45 and 3.52 phosphate buffers, the peak was found at 245 mμ and $\epsilon = 11{,}400$; in pH 5.40 buffer at 246 and $\epsilon = 10{,}400$; in pH 7.0 buffer at 248 mμ and $\epsilon = 8900$; in pH 8.8 at 248 mμ with $\epsilon = 8400$, and in pH 11.0 buffer at 260 mμ with $\epsilon = 3900$. Marasmic acid in the three most acid buffers had isobestic points at 230 and 255 mμ. There were no isobestic points at pH 7 and higher, indicating that a change in the light absorbing groups took place.

The positions of the peaks and the absorption coefficients is the same as that of several α,β-unsaturated ketones[7] in which the double bond is conjugate to the carbonyl group.

Marasmic acid reduced Fehling's reagent slowly in the cold and rapidly when heated. Ammoniacal silver nitrate was not reduced at room temperature but was reduced when boiled. When marasmic acid and hydroxylamine hydrochloride were mixed, there was no change in pH suggesting that the two substances did not react. Bromine in carbon tetrachloride did not add. Marasmic acid reduced cold potassium permanganate in sodium carbonate solution but not in cold acid solution. The acid did not give a color reaction with alcoholic ferric chloride, thus indicating that it probably is not a phenolic compound or an α-hydroxy acid. Marasmic acid reacted after a few seconds with alkaline nitroprusside to give a reddish color which became a very pale gray green when the solution was acidified. It formed an orange colored 2,4-dinitrophenylhydrazone with a corrected melting point of 136° to 138°C.

The reactions given above indicate that marasmic acid is unsaturated, contains a carbonyl group, and probably a carboxyl group, and changes at high pH so that increased reducing power results.

Marasmic acid added one molecule of cyanide for each equivalent of acid when the reaction (Kiliani) was conducted in 0.1 *N* acetic acid solution.[8] The 2,4-dinitrophenylhydrazones of marasmic acid and mesityl oxide had similar absorption bands in the near ultra-violet with the peak at 375 mμ when measured in alcohol. When the optical densities of the two dinitrophenylhydrazones were compared at 375 mμ on a molecular basis, the one from marasmic acid absorbed 1.12 times as much as the one from mesityl oxide, indicating that marasmic acid probably has one carbonyl group per equivalent weight.

A sample of marasmic acid was dissolved in four equivalents of 0.1 *N* sodium hydroxide, incubated at 37°C. for 2 hrs., and then back titrated with hydrochloric acid. The alkaline solution had a yellow color which disappeared as the solution became acid. There was no evidence for a second hydrogen, thus indicating the absence of a lactone ring.

Marasmic acid is not a methyl ketone or a $\beta,\gamma$-unsaturated lactone as judged by the slowness of the nitroprusside reaction.[9]

One equivalent of marasmic acid reacted with one molecular weight of cysteine as judged by the disappearance of the nitroprusside test for sulfhydryl. The reaction occurred within 2 hrs. at 37°C. and pH 6.

Marasmic acid was converted in alkaline solution in a short time into a product or products with chemical properties different from those of marasmic acid. The solution prepared by allowing marasmic acid to stand in about 0.1 *N* alkali for an hour and neutralizing, reacted immediately with nitroprusside to give a red color which faded to amber color within a minute or two. The 2,4-dinitrophenylhydrazone prepared from the solution did not crystallize from alcohol, chloroform, ether or benzene. The ultra-violet absorption spectrum of the alkaline-treated marasmic acid showed general absorption without a pronounced peak.

The reaction of marasmic acid with Fehling's solution, alkaline permanganate and nitroprusside, may depend upon conversion by alkali into the more reactive substance or substances referred to above.

On standing for a longer time in alkaline solution, the reactivity with nitroprusside was lost, indicating further change.

The behavior of marasmic acid suggests that it may be an $\alpha,\beta$-unsaturated ketone which shifts in alkaline solution so that the double bond is not conjugated with the carbonyl double bond.

*Antibacterial Activity of Marasmic Acid.*—The antibacterial and antiluminescent activities of the neutralized crystalline material were measured by the methods in use in this laboratory.[10] The results are given as the minimum concentration in micrograms per milliliter for inhibition of growth or luminescence.

ANTIBACTERIAL ACTIVITY

| | |
|---|---|
| *Bacillus mycoides* | 1 |
| *Bacillus subtilis* | 1 |
| *Escherichia coli* | 64 |
| *Klebsiella pneumoniae* | 128 (64 partial) |
| *Mycobacterium smegma* | 32 |
| *Photobacterium fischeri* | 0.016 |
| *Pseudomonas aeruginosa* | 250 (128 partial) |
| *Staphylococcus aureus* | 2 |

ANTILUMINESCENT ACTIVITY

| | |
|---|---|
| 10 minutes | 8 |
| 1 hr. | 4 |
| 2 hrs. | 2 |
| 3 hrs. | 1 |
| 24 hrs. | 0.25 |

*Effect of Blood.*—Marasmic acid was incubated with blood and the residual antibacterial activity determined for *Staph. aureus.* Human blood was diluted ten times in beef extract medium containing 0.7% sodium chloride. One milliliter of neutralized solution of marasmic acid was added to 1 ml. of the diluted blood, incubated for 3 hrs. at 37°C. centrifuged, and the antibacterial activity of the supernatant liquid determined by serial dilution. A control was diluted in beef extract medium without blood and incubated simultaneously. The activity of penicillin G, streptomycin and chloromycetin did not decrease on incubation with blood; the activity of marasmic acid was decreased. The greater the proportion of blood to marasmic acid, the greater was the decrease in activity.

*Antifungal Activity.*—The antifungal activity[11] of marasmic acid was measured by serial dilution in a peptone medium[12] at pH 6 using spore suspensions of the fungi. *Trichophyton* was incubated at 30°C.; the others at 25°C. The tests were read after from 42 to 48 hrs. incubation. The activity reported is the minimum inhibitory concentration in micrograms per milliliter; *p* = partial inhibition.

| ORGANISM | ACTIVITY |
|---|---|
| *Aspergillus niger* | 32 (16 *p*) |
| *Chaetomium globosum* (USDA 1042.4) | 64 |
| *Gliomastix convoluta* (PQMD4c) | 250 (8–128 *p*) |
| *Memnoniella echinata* (PQMD1c) | 64 |
| *Myrothecium verrucaria* (USDA 1334.2) | 64 |
| *Penicillium notatum* (832) | 32 (16 *p*) |
| *Phycomyces Blakesleeanus* (+ strain) | 16 (8 *p*) |
| *Saccharomyces cerevisiae* (188) | 32 |
| *Stemphylium consortiale* (PQMD41b) | 128 (64 *p*) |
| *Trichophyton mentagrophytes* | 4 |

*Animal Toxicity.*—Toxicity for mice was determined by injecting 0.5 ml. of neutralized marasmic acid made up in 0.7% sodium chloride solution into a tail vein of Carworth Farms CF1 male white mice. Each treatment group contained five mice. Four of the mice that received 32 mg./kg. in a single dose died within 3 days. All of the mice that received 16 mg./kg. were alive 30 days later. The LD50 lies somewhere between 16 and 32 mg./kg.

The high toxicity and the inactivation by blood militate against the possible therapeutic usefulness of marasmic acid.

*Additional Antibacterial Substances.*—The existence of additional antibacterial substances in the culture liquid of *Marasmius conigenus* was suggested by some of our observations.

The chloroform extracts of some lots of culture fluid were more active on a milligram basis than marasmic acid. For example, a spectrophotometric analysis of a chloroform extract indicated a maximum of 78% marasmic acid. The activity of this extract against *Staph. aureus* was 6 times that of an equal weight of crystalline marasmic acid. Five other strains of Staphylococci were from 4 to 32 times as sensitive to the material in the extract as to an equal concentration of marasmic acid. However, the activity for other bacteria tested was no different from that of marasmic acid. It appeared, therefore, that the culture liquid contained a factor to which Staphylococci were especially sensitive.

The antistaphylococcus factor was found in only four of 13 lots of culture liquid examined. Two of the lots of culture liquid assayed 64, one 256, and one 512 dilution units per milliliter against *Staph. aureus.* The concentrates containing the antistaphylococcus factor were all prepared from solutions in which the fungus had been grown from the inoculum; the factor was not observed in solutions from flasks that had been reflooded. The antistaphylococcus factor was destroyed by heating a dry fraction in an oven for 3 hrs. at 100°C. Marasmic acid was unaffected by this treatment. The antistaphylococcus factor was not affected by air drying overnight nor by bringing it to a boil in a solution of pH 2 to 3. Incubation with blood reduced the activity of the antistaphylococcus factor.

A further series of observations suggested that *Marasmius conigenus* may produce a third antibacterial substance. Chloroform removed all or nearly all the antibacterial activity from the various lots of acidified culture liquid as indicated by tests on the aqueous phase after extraction. However, the proportion of original activity of the culture liquid present in the chloroform extract varied with different lots from 3 to 100%. We did not determine the cause for this variability. It seems improbable that loss or destruction of either marasmic acid or the antistaphylococcus factor accounts for the variability in recovery of antibacterial activity in the chloroform extracts, and, therefore, a third less stable substance may have been present.

---

* This investigation was supported in part by grants from The Commonwealth Fund and the Albert H. & Jessie D. Wiggin Foundation.

[1] Wilkins, W. H., *Trans. Brit. Mycol. Soc.*, **28**, 110–114 (1945).

[2] Melin, E., Wikén, T., Öblom, K., *Nature*, **159**, 840–841 (1947).

[3] Bendz, G., Wallmark, G., Öblom, K., *Ibid.*, **162**, 61–62 (1948).

[4] Robbins, W. J., Hervey, A., Davidson, R. W., Ma, R., and Robbins, W. C., *Bull. Torrey Bot. Club*, **72**, 165–190 (1945).

[5] Hervey, A., *Ibid.*, **74**, 476–503 (1947).

[6] The analysis was made by Mr. Joseph F. Alicino.

[7] Rasmussen, R. S., Tunnicliff, D. D., and Brattain, R. R., *Jour. Am. Chem. Soc.*, **71**, 1068–1072 (1949).

[8] Militzer, W. E., *Archiv. Biochem.*, **9**, 91–94 (1946).

[9] Jacobs, W. A., Hoffman, A., and Gustus, L., *Jour. Biol. Chem.*, **70**, 1–11 (1926).

[10] Kavanagh, F., *Bull. Torrey Bot. Club*, **74**, 303–320, 414–425 (1947).

[11] We are indebted to Dr. W. L. White for the cultures with the USDA or PQMD numbers, to Dr. R. F. Light, Fleischmann Laboratories, for the culture of *S. cerevisiae*.

[12] This medium contained per liter 1.5 g. $KH_2PO_4$, 0.5 g. $MgSO_4 \cdot 7H_2O$, 50 g. dextrose, 2 g. neopeptone, 1500 mμ moles of thiamine, and a mixture of minor mineral elements as used in this laboratory.

---

# *CHROMOSOME SEGREGATION IN MAIZE TRANSLOCATIONS IN RELATION TO CROSSING OVER IN INTERSTITIAL SEGMENTS**

BY C. R. BURNHAM†

CALIFORNIA INSTITUTE OF TECHNOLOGY AND THE UNIVERSITY OF MINNESOTA‡

Communicated by G. W. Beadle, May 23, 1949

In maize, the second meiotic division in the microsporocytes results in a quartet of four microspores in which it is possible to distinguish the two division planes. In normal material, the prominent feature of these spores, as revealed by acetocarmine smears, is the single nucleolus. In translocations involving chromosome 6 which carries a nucleolar organizer region, non-disjunction may result in spores with two nucleolar organizers (potentially two nucleoli), or with no organizer (nucleolar material remains scattered or diffuse) in the same quartet. By this method, McClintock[4] established the fact that the chromosomes which crossover in an interstitial segment (between the centromere and the translocation break) pass to opposite poles.

A further relation between chromosome segregation in such maize translocations and the frequency of crossing over in an interstitial segment has been reported in abstracts.[1, 2] This is a report of additional studies of

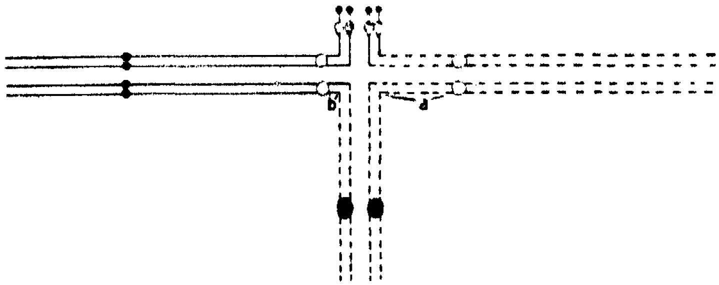


FIGURE 1

Pachytene diagram of a 5-6 translocation heterozygote with the break in chromosome 6 (solid line) in the short arm, that in 5 (dotted line) in the long arm, *b* and *a* being the interstitial segments in the respective chromosomes. The satellite and next to it the nucleolar organizer region (stippled) are shown at the end of the short arm of chromosome 6. The clear circles represent the centromeres, the dense ones the pycnotic knobs.

TABLE 1

SPORE QUARTET TYPES FROM A TRANSLOCATION HETEROZYGOTE WITH THE BREAK IN THE SHORT ARM OF CHROMOMOSOME 6

The figures 0, 1, 2 in the body of the table refer to the quartets containing no, one, and two spores with diffuse nucleolar material. The "no diffuse" may have four normal or four abortive spores, the latter indicated as (ab) in the table; the "two diffuse" have four abortive, and the "one diffuse" has two normal and two abortive, except a few cases where all four abort, indicated by (ab).

| EVENTS IN INTERSTITIAL SEGMENTS | ORGANIZER MAKE-UP OF CHROMATID PAIR | SPORE QUARTET TYPES AFTER SEGREGATIONS OF ALTERNATE | ADJACENT 1 | ADJACENT 2 |
|---|---|---|---|---|
| No crossover in *a* or *b* | $\frac{\cdot\ \cdot}{-\ -}\quad\frac{\cdot\ \cdot}{-\ -}$ | 0 | 2 | 0 (ab) |
| Single crossover in *a* or *b* | $\frac{\cdot\ \cdot}{-\ -}\quad\frac{\cdot\ -}{\cdot\ -}$* | 1 | 1 | 0 (ab), 1 (ab) |
| Single crossovers in *a* and *b* | $\frac{\cdot\ -}{\cdot\ -}\quad\frac{\cdot\ -}{\cdot\ -}$* | 0, 2, 1† | 0, 2, 1† | 0 (ab), 2, 1† |
| Double crossover in *a* or *b* | | | | |
| 2-strand | $\frac{\cdot\ \cdot}{-\ -}\quad\frac{\cdot\ \cdot}{-\ -}$ | 0 | 2 | 0 (ab) |
| 3-strand | $\frac{\cdot\ \cdot}{-\ -}\quad\frac{\cdot\ -}{\cdot\ -}$ | 1 | 1 | 0 (ab), 1 (ab) |
| 4-strand | $\frac{\cdot\ \cdot}{-\ -}\quad\frac{-\ -}{\cdot\ \cdot\ \cdot}$ | 2 | 0 | 0 (ab) |

* Crossover chromatids, see note at end of paper. . = organizer, - = its absence.

† Ratio of 1:1:2 in each. Here in the two-diffuse type the two diffuse spores are on opposite sides of the division I plane, and in the one-diffuse type the two abnormal spores are on the same side. In the other lines of the table, the two diffuse spores are on the same side and in the one-diffuse quartet the two abnormal ones are on opposite sides of the division I plane.

chromosome segregation in maize translocations involving chromosome 6, the detailed data to be published elsewhere. The ultimate objective is to determine why chromosome segregation is directed (i.e., largely alternate) in certain species and not in others.

The results differ, depending on whether the break in chromosome 6 is in the long arm or in the short arm between the centromere and the nucleolar organizer. The theoretical results for a translocation with the break in 6 in the short arm will be considered first, followed by the experimental data.

✓The pachytene configuration in a translocation heterozygote with one break in the short arm of 6 between the centromere and the nucleolar organizer is shown in Fig. 1. Alternate or adjacent chromosomes in this complex may pass to the same pole, the latter being of two types—adjacent 1, in which homologous centromeres pass to opposite poles at division I, and adjacent 2 in which they pass to the same pole. When no crossing over has occurred in the interstitial segment, adjacent 1 brings about non-disjunction of the translocated pieces including the nucleolar organizers; adjacent 2 brings about non-disjunction of the non-translocated pieces. Unless stated otherwise, the crossovers referred to in this paper are the cytological ones occurring in interstitial segments. Crossovers in other segments do not change the quartet types or the pollen abortion.

The theoretically expected spore quartet types with and without crossing over in the interstitial segments are listed in table 1. The second column of this table shows the nucleolar organizer constitution of the chromatid pairs (sister centromeres) after crossing over, i.e., before orientation. Reference to this column should facilitate determining the quartet type resulting from each kind of segregation. With no crossing-over, alternate as well as adjacent 2 segregation results in a "no diffuse" spore quartet, the former with four normal or functional spores, the latter with four abortive ones, the two being indistinguishable cytologically, while adjacent 1 segregation results in the "two-diffuse" type in which the four spores are expected to be abortive (two spores have potentially two nucleoli, two have diffuse nucleolar material). A single crossover or a 3-strand double followed either by alternate or by adjacent 1 segregation results in the "one diffuse," also termed a "crossover-type" quartet, expected to have two functional (one nucleolus each) and two abortive (one with potentially two nucleoli, one with diffuse nucleolar material) spores. Adjacent 2 segregation following these crossovers is not expected since the evidence indicates that chromosomes that cross over pass to opposite poles. If it did occur, a "no-diffuse" and a "one-diffuse" type of spore quartet would be expected, but the two abnormal spores in the latter are on the same side of the first division plane. The absence of this class in a translocation with only one interstitial segment furnished the evidence that chromosomes that cross over pass to opposite poles.[4] Simultaneous crossing over in both interstitial segments gives rise to the three quartet types following each of the three kinds of segregation (table 1). Consequently, to determine segregation types by this method it is necessary to use translocations in which these simultaneous crossovers are infrequent, i.e., at least one segment genetically short, or if both are long, to have a genetic measure of the amount of simultaneous crossing over.

The three available translocations in which one break was in the short arm of chromosome 6, were: T2-6a (2L.4?, 6S?), T6-10b (6S.5, 10L.58), and T5-6c (5L.89, 6S0.0). In T5-6c the break in 6 was adjacent to the centromere, so that there is no interstitial segment in chromosome 6, while in the other two translocations it is not long enough to give much crossing over since this arm of 6 is very short. The break in T2-6a is in the short arm but the exact point was not determined. The interstitial segment in the other chromosome of these translocations differed in length. An inversion in chromosome 5 (In5a-S0.0-L.67) served to shift the position of the centromere in chromosome 5. By crossing a stock homozygous for T5-6c and In5a with one homozygous only for In5a, plants were produced which were homozygous for In5a and heterozygous for T5-6c. In these plants, due to the shift in centromere position, the interstitial segment in chromosome 5 is short (a length equal to 0.22 of the long arm). In those heterozygous for T5-6c and not carrying the inversion that segment is long (0.89 of the long arm). In both T5-6c/+ stocks the relative lengths of the two axes and the lengths of the translocated pieces remain the same.

TABLE 2

SUMMARY OF SPORE QUARTET AND POLLEN STERILITY DATA AND FREQUENCIES OF CHROMOSOME SEGREGATION TYPES IN PLANTS HETEROZYGOUS FOR TRANSLOCATIONS HAVING ONE BREAK IN THE SHORT ARM OF CHROMOSOME 6; THE FIRST THREE WITH SHORT INTERSTITIAL SEGMENTS, THE LAST WITH ONE VERY LONG.

| | QUARTETS | | | | POLLEN ABORTION % | | | CHROMOSOME SEGREGATION IN NON CROSSOVER QUARTETS | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | % | | | OBSERVED | | | | | |
| TRANSLOCATION HETEROZYGOTE | TOTAL | NO DIFFUSE | TWO DIFFUSE | ONE DIFFUSE | TOTAL | PARTLY FILLED | PREDICTED | ALTERNATE | ADJACENT 1 | ADJACENT 2 |
| T5-6c In5a / + In5a | 7189 | 61.7 | 27.8 | 10.6 | 50.5 | 13.1 | 33.0 | 49.4 | 31.0 | 19.6 |
| T2-6a / + | 2361 | 73.6 | 25.3 | 1.1 | 51.6 | .. | 25.7 | 48.4 | 26.2 | 25.5 |
| T6-10b / + | 2374 | 76.6 | 18.1 | 5.3 | 42.9 | .. | 20.7 | 58.0 | 22.4 | 19.7 |
| T5-6c / + | 4229 | 20.3 | 17.9 | 62.8 | 48.1 | 22.1 | 49.3 | ..* | ..* | 0.0 |

* These classes cannot be determined.

The frequencies of the different spore quartet types were determined cytologically (in all but the early studies, without reference to division planes) and from them the expected degree of pollen abortion was calculated. The amount by which the observed pollen abortion exceeded the calculated was used as a measure of the frequency of adjacent 2 segregation. The data for each translocation are based on at least two plants and two slides for each, each slide made from one anther from different florets. The counts on the pollen to determine the percentages of abortion were made on the same plants. The spore quartet and pollen abortion data and the calculated segregation frequencies for these translocations are summarized in table 2. Comparison shows the first three stocks have 20 to 26% of adjacent 2 segregation while the last one, T5-6c/+, has none. In the latter, the observed pollen abortion was actually higher than that predicted from the cytological count of the spore quartets. In the first three, the percentages of crossover-type quartets were low, 10.6, 5.3 and 1.1, respectively, while in the latter it was high, 62.8%. This indicates a close relation between the length of the interstitial segment, in terms of cytological crossing-over, and chromosome segregation, i.e., low crossover is accompanied by frequent adjacent 2, high crossover is accompanied by little or no adjacent 2 segregation. In the stock with 62.8% of crossover-type quartets, there was probably a considerable amount of double crossing-over and some higher multiples. If as much as 15% doubles occurred, practically all the tetrads would have had one or more crossovers in the interstitial segment, using Weinstein's[6, 7] and Sax's[5] method of calculation. The absence of adjacent 2 segregation is then expected since chromosomes that cross over pass to opposite poles. Adjacent 2 segregation could occur only in those with no such crossovers. A comparison of the results for the two T5-6c/+ stocks is especially convincing since in these the lengths of the translocated pieces remained the same, the difference being in the length of the interstitial segment. The differences in segregation, 20% of adjacent 2 in one, none in the other, were achieved in spite of very unequal axes lengths and no change in the lengths of axes or of the translocated pieces. In an earlier report[1] the differences were supposed to be the result of a centromere effect. It now seems probable that they are the result of the relation between crossing-over in the interstitial segment and disjunction, centromere position being incidental in determining the length of that segment, but still important in that it may initiate the separation of the crossover chromosomes. Although probable in corn, there is no proof in this study that alternate and adjacent 1 segregations following these crossovers are equally frequent, hence the frequencies of alternate and adjacent 1 segregations are not calculated for T5-6c/+ in table 2.

In those translocations with a low frequency of crossover-type quartets, the two-diffuse class represents the frequency of adjacent 1 segregation since double crossovers probably were lacking. The frequency of quartets from alternate segregation is the difference between the total of the no-diffuse class and those from adjacent 2. From the segregation frequencies calculated on this basis, for the first three stocks in table 2 it will be noted that alternate segregation is roughly 50%, adjacent 1 and 2 segregations making up the other 50%. In other words, adjacent 2 segregation is probably at the expense of adjacent 1.

Translocations with one break in the long arm of chromosome 6 were also studied. In these, only two quartet types can be recognized cytologically, no-diffuse and two-diffuse, the latter following adjacent 2 segregation with or without crossing-over in the interstitial segments. Here when there is little crossing-over, the difference between observed and predicted pollen abortion measures the frequency of adjacent 1 segregation. Segregation of three chromosomes of the ring to one pole and one to the other, resulting in $n + 1$ and $n - 1$ spores, also are expected (part of them) to result in the two-diffuse spore quartet, offering some difficulty. Usually its total frequency is not expected to be higher than about 5%. For ten such translocations with short interstitial segments, adjacent 2 segregation ranged from 14 to 36%, the average for the entire group being 25%. Where counts could be made, pollen abortion was about 50% in this group, this representing the total of the two types of adjacents. For eight translocations in which either or both interstitial segments were probably long enough for a high frequency of crossing-over (breaks at 0.6 or greater in the longer chromosomes and 0.7 or greater in the shorter ones), the average was 5.4% of adjacent 2 segregation. One, T6-8a with 17.4%, appears to be an exception. However, another T6-8, (D1), with breaks at about the same points had only 1.1% of adjacent 2. Some unknown factor seems to have been operating in T6-8a. With this one exception, the relationship appears to be that when the interstitial segments are short, both types of adjacents occur with relatively high frequencies, not always equal but their total is roughly 50%. When one or both interstitial segments are long, adjacent 2 segregation is low in frequency.

Chromosome segregation was studied also in plants heterozygous for translocations having one break in the satellite of chromosome 6 and which, therefore, always form a chain of four chromosomes. Two were studied which have short interstitial segments, T1-6b and T5-6b, and one, T3-6b, with one long interstitial segment. In the first two there was less than 1% of adjacent 2 segregation, and in the latter none. This result, therefore, differs from that for the ring-forming translocations since in them, those with short interstitial segments had a relatively high frequency of adjacent 2 segregation. In T5-6c/+ which has one long interstitial segment, 30.6% of chains were observed at diakinesis, and no measurable adjacent 2 segregation. The evidence from these translocation heterozygotes indicates that adjacent 2 segregation from chains is infrequent whether the amount of crossing-over in the interstitial segment is low or high. Adjacent 2 segregation, therefore, occurs rarely except in rings in which no crossing-over has occurred in the interstitial segments.

*Discussion and Summary.*—In maize translocation heterozygotes which form rings and have short interstitial segments (low crossover frequency), alternate segregation occurs in about 50% of the meiocytes. In them, both types of adjacent segregations occur, not always with equal frequency but the total is roughly 50%. In those with interstitial segments long enough to permit a high crossover frequency there is little or no adjacent 2 segregation. The segregation in rings differs greatly from that in chains. In chains, there is little or no adjacent 2 segregation whether the interstitial segment is long or short (in terms of crossing-over).

There is, therefore, no evidence of any directed segregation in the maize translocations studied, even with greatly different break positions. This, together with Catcheside's observations[8] that segregation is directed even in unequal translocations produced by x-raying an *Oenothera* race with seven pairs suggest that directed segregation may be genetically controlled. Even in a species homozygous for such a genotype, translocations with long interstitial segments should have considerable sterility as a result of crossing-over in those segments, reaching 50% as a possible maximum. According to the literature, the translocations that have survived in *Oenothera* are those with short interstitial segments. These, therefore, would have low sterility from that source. A search for genes in maize affecting segregation is being made by crossing stocks of widely different origin with translocations having short interstitial segments.

*Acknowledgment.*—I wish to express my gratitude to Dr. Barbara McClintock who originally suggested the problem and furnished the stocks of T5-6c, T6-10b, In5a, and the stock combining T5-6c with In5a. Throughout the course of the experiments, assistance has been furnished from research funds of the University of Minnesota graduate school, for which I am deeply grateful. A sabbatical leave, together with a Gosney Research Fellowship at the California Institute of Technology (1947–1948), made its completion possible. I am indebted to Dr. E. G. Anderson for his generous coöperation. I wish to acknowledge the assistance of Gertrud Joachim, C. H. Li and H. R. Highkin in gathering the pollen sterility and spore quartet data.

[*] Published as Paper No. 2475 of the Scientific Journal Series, Minnesota Agricultural Experiment Station, St. Paul, Minn.

[†] Gosney Fellow, California Institute of Technology.

[‡] Permanent address.

NOTE: Lamm, *Hereditas*, 34, 280–288 (1948) (Dr. H. H. Kramer called my attention to this paper as mine was being completed), and Kramer and Hanson (*Genetics*, in press) have pointed out that chromatids resulting from single crossovers in the interstitial segment are recovered only when adjacent 1 segregation occurs, and are lost following alternate segregation. This may be seen by referring to column 2 of table 1 in which the crossover chromatids are marked with an asterisk. The results of multiple crossovers are not indicated. In species with more alternate than adjacent 1 segregation, genetic crossing-over in this region will be reduced. My conclusions in this paper are based on frequencies of cytological crossing-over, and consequently are not altered by these relations.

[1] Burnham, C. R., *Genetics*, **30**, 2 (abstract) (1945).

[2] Burnham, C. R., *Ibid.*, **33**, 605 (abstract) (1948).

[3] Catcheside, D. G., *Genetica*, **17**, 313–341 (1945).

[4] McClintock, B., *Am. J. Bot.*, **32**, 671–678 (1945).

[5] Sax, K., *J. Arnold Arboretum*, **13**, 180–212 (1932).

[6] Weinstein, A., *Anat. Rec.*, **41**, 109–110 (1928).

[7] Weinstein, A., *Genetics*, **21**, 155–199 (1936).

---

# *EARLY TERTIARY ECOTONES IN WESTERN NORTH AMERICA*

By Ralph W. Chaney

Department of Paleontology, University of California,
and Carnegie Institution of Washington

Read before the Academy, April 25, 1949

Within recent years we have come to realize that the concept of cosmopolitan land floras during past ages is not substantiated by the fossil record of the Tertiary period. Similar floras may be recognized over a wide range in latitude, but differences in their age are always apparent. We now interpret their shifting positions during later geologic time as resulting from migrations largely induced by changes in climate. Varying distance from the equator seems always to have involved sufficiently great differences in temperature to leave its impress on land vegetation. During periods of submergence, when there was wide circulation of winds and ocean currents, more uniform climates appear to have characterized broad areas of the earth; but even at such times the fossil record shows marked differences in forest composition from south to north.

Major units of living vegetation, although largely confined to latitudinal zones, show at their borders a tendency to merge into one another. We may draw forest boundaries on a map with a firm hand, but in nature such lines are commonly blurred. A similar overlapping of floras may be observed in the records of past vegetation. Transitional occurrences of this sort, past and present, are known as ecotones; since they are largely a function of temperature they extend along the parallels, bending southward along mountain ranges in the northern hemisphere, and north on windward shores. Ecotones provide a critical basis for determining the tolerance range of species and groups of plants. When they are recognized in vegetation of the past they provide much-needed information regarding former environments; this is particularly the case with Tertiary floras, in which the general composition of modern plant groups was already well established, permitting an accurate comparison with existing vegetation. I shall consider here three of the ecotones in the older Tertiary of western North America which have been reviewed recently in connection with my studies of Metasequoia. Other ecotones of this age and younger are available for study in many parts of the northern hemisphere.

The older Tertiary floras of western North America may be assigned to two groups on the basis of their climatic indications. The first of these is a subtropical assemblage made up predominantly of broad-leafed *evergreen* angiosperms whose living equivalents occupy regions in which freezing temperatures are seldom if ever recorded; palms and cycads are represented by a few species which may be abundant; conifers are rare or absent. This forest occupied a zone from California north to Washington on the Pacific Coast, and from New Mexico to Colorado in the interior. Because of its apparent origin in low American latitudes, it is known as the Neotropical-Tertiary Flora. The second major group is a temperate assemblage made up predominantly of broad-leafed *deciduous* angiosperms, and of conifers both deciduous and evergreen; palms and cycads have not been recorded. This forest occupied a zone in Alaska northward beyond the Arctic Circle, extending into the Mackenzie Basin of Canada. Evidence of distribution farther eastward has been destroyed by Pleistocene glaciation, but its former occurrence across northern Canada is suggested by the records of similar plant fossils in the Eocene rocks of Greenland, Spitzbergen and arctic Siberia. Because of its wide development at high northern latitudes, this assemblage is known as the Arcto-Tertiary Flora. It lived in a region of summer rainfall and winter cold, as judged by the habits of related living trees.

Two aspects of early Tertiary zoning may here be noted. In the first place, the zones were located approximately 20 degrees north of areas where vegetation of similar kinds is now found; it has been concluded that climate at the beginning of the Tertiary period was characterized by much higher temperatures than those prevailing today at corresponding latitudes, with an absence of severe frost as far north as northern Washington (48° N. Lat.).[1] Second, the zones extend to a higher latitude on the Pacific border than in the interior; this is interpreted as indicating that as far back as Eocene time the Pacific shore was at approximately its present position, with temperatures moderated by the ocean along the coast, and with continental climate at the interior.[2]

In the areas between these two major zones of vegetation there is a record of a forest made up largely of temperate trees. Living with them were several members of the subtropical forest which appear to have made their way northward beyond the range of occurrence of most of their associates, during the relatively mild days of the early Tertiary. This ecotone is best developed in southeastern Alaska, on Kupreanof Island at latitude 56 to 57 degrees north. Here is the typical Kenai flora, as described by Hollick from 55 localities in Alaska, including such temperate trees as redwood (Metasequoia), katsura (Cercidiphyllum), birch (Betula), alder (Alnus), hazel (Corylus), hornbeam (Carpinus), chestnut (Castanea), oak (Quercus) and elm (Ulmus).[3] Immediately associated, and often on the same slabs of rock, are the leaf impressions of several members of the Neotropical-Tertiary Flora; these include two genera of cycads (Dioon and Ceratozamia), a fan palm (Flabellaria), laurels of the avocado (Persea) and lancewood (Ocotea) type, magnolia (Magnolia) and catmón (Dillenia). This is the farthest north occurrence of the Neotropical-Tertiary Flora.

In the Rocky Mountains of southern Saskatchewan, and in Montana and North Dakota, there is a similar mingling of northern with southern trees. Oak, chestnut, elm and birch occur with cycads, palm, magnolia, fig and laurels. Metasequoia is the common conifer of this forest of the Fort Union formation, with an age somewhat older than the Alaska flora.[4]

The modern occurrence of Metasequoia in a restricted area in central China shows a striking resemblance to these ecotones of the early Tertiary.[5] Here Metasequoia is growing in immediate association with birch, chestnut, katsura and other Arcto-Tertiary genera, as well as with palm, magnolia and laurels, trees commonly found in the Neotropical-Tertiary and Paleotropical-Tertiary Floras. The similarity of this living forest to the fossil floras of southeastern Alaska, Saskatchewan and Montana is so close as to demonstrate the survival of this subtropical-temperate ecotone for several scores of million years with only slight modification. There is one significant difference: the modern Metasequoia ecotone shows an *altitudinal* relationship, between the subtropical trees at lower levels and the temperate trees of the highlands; the Metasequoia ecotones of the early Tertiary show a *latitudinal* relationship between trees of subtropical and temperate zones.

Since the fossil record rarely includes the remains of upland forests, altitudinal ecotones like that of central China will not often be encountered. For a number of years we have known of a mixture of subtropical with temperate trees in the Eocene Clarno flora of the John Day Basin of Oregon.[6] Here, in a region whose older Tertiary floras are typically of a subtropical type, has been found a florule combining birch, elm and oak with Mallotus, Cedrela and other trees now limited to low latitudes. This is the best known fossil occurrence of an altitudinal ecotone, and again its most conspicuous member is Metasequoia.

We may conclude from the plant record of the Tertiary period that

Metasequoia has long been a member of the forest of the temperate zone, and has commonly ranged southward into latitudinal ecotones with the forests of the subtropical zone. During the Eocene epoch this conifer extended still farther south in the uplands, where it lived in an altitudinal ecotone bordering the Clarno flora of the John Day Basin. It has survived in the high valleys of central China under conditions which seem to resemble closely its Tertiary environment in Oregon. The modern occurrence of *Metasequoia glyptostroboides* in the mixed deciduous and evergreen forest of Szechuan and Hupeh therefore becomes a key to our understanding of many forests of the past in which this deciduous conifer was a member of similar ecotones.

[1] Chaney, R. W., *Ecol. Mon.*, **17**, 144, 146 (1947).

[2] Chaney, R. W., *Bull. Geological Survey Soc. Am.*, **51**, 481–486 (1940).

[3] Hollick, A., *U. S. Geol. Survey, Prof. Paper* **182** (1936).

[4] Study of the flora of the Fort Union formation is now being completed by R. W. Brown of the U. S. Geological Survey, to whom acknowledgment is due for generic identifications.

[5] Chaney, R. W., these PROCEEDINGS, **34**, 503–515 (1948).

[6] Chaney, R. W., *Carnegie Inst. Wash. Pub.*, **501**, 636 (1938).

---

# *THE DISSOCIATION ENERGY OF CARBON MONOXIDE AND THE HEAT OF SUBLIMATION OF GRAPHITE*

BY LINUS PAULING AND WILLIAM F. SHEEHAN, JR.

GATES AND CRELLIN LABORATORIES OF CHEMISTRY, CALIFORNIA INSTITUTE OF TECHNOLOGY*

Communicated May 5, 1949

For a number of years there has existed doubt about the value of the dissociation energy of carbon monoxide and about the heat of sublimation of graphite, a directly related quantity. The most popular values for the dissociation energy of carbon monoxide are 9.144 electron-volts, suggested by Herzberg[1] on the basis of predissociation phenomena in band spectra, 9.61 e. v., suggested by Hagstrum and Tate[2] on the basis of electron impact experiments (or the value 9.85 e. v. derivable from predissociation data[3]), and 11.11 e. v., suggested by Gaydon and Penney[4] from an analysis of spectroscopic data. These values together with thermochemical data lead to the values 124.9, 141.4, and 170.3 kcal./mole, respectively, for the heat of sublimation of graphite. Strong evidence for the last of these values has been presented by Brewer, Gilles, and Jenkins,[4] who have reported 170.4 kcal./mole from a direct experimental determination. The value has, however, been criticized by other investigators,[6, 7] and has been defended by Brewer.[8]

In this paper we communicate an argument which indicates that the high values of about 170 kcal./mole for the heat of sublimation of carbon to $C_{(g)}(^3P)$ and 11.11 e. v. for the dissociation energy of carbon monoxide are not correct, and which leads instead to the values 140 kcal./mole and 9.77 e. v., respectively.

The argument is based on the postulate that the linear extrapolation of the vibrational levels for the lower vibrational states of a molecule (a linear Birge-Sponer extrapolation) leads to an energy value corresponding to the dissociated atoms in a hypothetical electronic state called the valence state. An atom in the valence state, as defined in this paper, is an isolated atom with the same electronic structure as that which the atom has in the molecule under consideration. It is the state that would result if the atoms in the molecule were to be pulled apart without change in the electronic structures that exist in the molecule in the lower vibrational levels of the lower electronic state. The wave function for the valence state of an atom may of course be formed by linear combination of those for its spectroscopic states, and the energy of the valence state is somewhat higher than that of the normal spectroscopic state, by an amount, the valence-state energy, that is reasonably constant for a given atom from molecule to molecule.[9, 10]

The valence-state energy of the oxygen atom has been evaluated as 0.74 ± 0.05 e. v.[10] A rough value for the valence-state energy of nitrogen can be calculated from spectroscopic data for the three low-energy levels of the atom, all based on the configuration $2s^22p^3$. These levels, $^4S$, $^2D$, and $^2P$, have energy values $F^\circ$, $F^\circ - 6/25\,F^2$, and $F^\circ - 15/25\,F^2$, respectively, according to simple spectroscopic theory. The states $^2D$ and $^2P$ are observed to lie at energies 2.38 e. v. and 3.57 e. v. above the normal state $^4S$. These values are not in the ratio 3:5 given by the simple theory, and accordingly there is uncertainty as to the value of the resonance integral $F^2$. The normal valence state of nitrogen to the extent that it is based on the normal configuration $2s^22p^3$ is the state in which each of three electrons occupying separate $p$ orbitals has its spin oriented independently of the other two electrons, the corresponding energy being $F^\circ - 21/50\,F^2$. Because of the uncertainty in the value of $F^2$, the energy of the valence state cannot be predicted precisely. If the $^4S$ level is considered to be depressed by resonance with a similar state based on an excited configuration of the atom the valence-state energy would be calculated to be 1.39 e. v., whereas if the $^2P$ level is considered to be depressed by resonance the valence-state energy would be 1.19 e. v. A value somewhat larger than either of these values might be expected to result from contributions of higher spectroscopic states to the valence state.

A more reliable value for the valence-state energy of the nitrogen atom can be obtained from the consideration of nitric oxide. The energy of dissociation of the normal nitric oxide molecule to a nitrogen atom and an oxygen atom in their valence states is found by linear extrapolation of the low-lying vibrational levels to be 7.87 e. v. If we subtract from this quantity the sum of the valence-state energy of oxygen, 0.74 e. v., and that of nitrogen, approximately 1.5 e. v., the dissociation energy $D_0$ is predicted to be approximately 5.6 e. v. This argument hence favors the value $D_0 = 5.29$ e. v. proposed by Mulliken[11] and supported by Wulf[12] and Hagstrum,[13] rather than the value 6.49 e. v. supported by Gaydon.[14] If we accept the value 5.29, the valence-state energy of the nitrogen atom is calculated to be 1.84 e. v.

Some substantiation of this value is provided by the consideration of the normal state and the first excited state of the nitrogen molecule. The energy of dissociation of $N_2$ in its normal state $X^1\Sigma_g^+$ into two nitrogen atoms in their valence states is found by linear extrapolation to be 11.60 e. v., and the value of this quantity given by the first excited state $A^3\Sigma_u^+$ is 10.58 e. v. On subtracting the dissociation energy of the nitrogen molecule, 7.38 e. v. (the alternative spectroscopic value 9.76 e. v. is to be eliminated as not leading to sufficiently large values for the valence-state energy and being incompatible with the results of electron-impact experiments[13]), and dividing by 2, we obtain 2.11 and 1.60 e. v., respectively, for the valence-state energy of the nitrogen atom. The value 2.11 e. v. given by the normal state of the nitrogen molecule is probably high because of a large amount of $s$ character in the orbital of the $\sigma$ bond for this molecule. We conclude that the normal valence-state energy of the nitrogen atom is approximately 1.84 e. v., and that variation of a few tenths of an electron-volt may be expected.

The valence-state energy of the carbon atom can be similarly derived from spectroscopic data for the molecules CH and $C_2$. For the normal state of CH the energy of dissociation to a carbon atom and a hydrogen atom in their valence states is 5.20 e. v., as given by linear extrapolation, and the dissociation energy $D_0$ to the atoms in their normal states is 3.47 e. v. The difference, 1.73 e. v., can be taken as the valence-state energy of the carbon atom, inasmuch as the valence state and the normal spectroscopic state of the hydrogen atom are essentially the same. For the normal $C_2$ molecule the energy of dissociation to two carbon atoms in their valence states is found by linear extrapolation to be 7.05 e. v., and the dissociation energy $D_0$ is 3.6 e. v. These quantities lead to 1.73 e. v. for the valence-state energy of carbon, in exact agreement with the preceding value.

The energy of a carbon atom in its valence state and an oxygen atom in its valence state relative to the normal state of the carbon monoxide molecule is found to be 11.23 e. v. by linear extrapolation for the normal state, and the values 12.33, 11.87, and 12.10 e. v. are similarly obtained for the first three excited states of the molecule, $a\ ^3\Pi$, $a'\ ^3\Sigma$, and $A\ ^1\Pi$. If we subtract from these numbers the quantity 2.47 e. v., the sum of the valence-state energies for carbon and oxygen, as evaluated above, we obtain 8.76, 9.86, 9.40 and 9.63 e. v. for the dissociation energy of the normal carbon monoxide molecule to atoms in their normal spectroscopic states. These values are incompatible with the high value 11.11 e. v. for $D_0$(CO), and may be considered to support either the value 9.14 or the value 9.61–9.85.

Evidence indicating that the second of these values, 9.6–9.85, is correct is obtained from the discussion of the cyanogen molecule and the cyanide radical. Linear extrapolation of the vibrational levels of the lowest state of the CN molecule, $X\ ^2\Sigma^+$, leads to 9.85 e. v. for the energy of a carbon atom and a nitrogen atom in their valence states relative to the normal state of the molecule. On subtracting 1.73 for the valence-state energy of carbon and 1.84 for that of nitrogen, we obtain 6.28 e. v. for $D_0$(CN). This value strongly supports the spectroscopic value 6.24 e. v. for this molecule,[14] rather than the alternative spectroscopic value 7.50 e. v. From thermochemical data and the accepted dissociation energy of the oxygen molecule, 5.08 e. v., it can be calculated[15] that the dissociation energy of cyanogen, $C_2N_2$, into two CN molecules is given by the equation

$$D_0(\mathrm{NC{-}CN}) = 2D_0(\mathrm{CO}) + D_0(\mathrm{N_2}) - 2D_0(\mathrm{CN}) - 10.7 \text{ e. v.} \quad (1)$$

The dissociation energy of cyanogen has been reported by Kistiakowsky and Gershinowitz[16] to be 77 ± 4 kcal./mole, and by White[17] to be 146 ± 4 kcal./mole. A chemical argument can be given that leads to an intermediate value. The dissociation energy of cyanogen may be expected to be greater than the carbon-carbon single-bond energy in diamond (one-half the heat of sublimation of diamond) by an amount equal to the conjugation energy of the two triple bonds in the molecule. This conjugation energy is twice that of two conjugated double bonds, with theoretical value[18] 8 kcal./mole. The value of $D_0$(NC—CN) may hence be taken as 16 kcal./mole (0.70 e. v.) greater than one-half the heat of sublimation of diamond, and if this heat of sublimation lies between 125 and 170 kcal./mole the value of $D_0$(NC—CN) is calculated to be 90 ± 11 kcal./mole Instead of using this value, with its large uncertainty, we may adopt the following procedure. Thermochemical data require that the value of $D_0$(CO) be 3.70 e. v. greater than the heat of sublimation of diamond, and hence, with the resonance energy 0.70 e. v. in cyanogen, we obtain the equation

$$D_0(\mathrm{NC{-}CN}) = {}^1/_2 D_0(\mathrm{CO}) - 1.15 \text{ e. v.} \quad (2)$$

From (1) and (2) we obtain the equation

$$3/2\, D_0(\mathrm{CO}) = 2D_0(\mathrm{CN}) - D_0(\mathrm{N_2}) + 9.55 \text{ e. v.} \quad (3)$$

This equation with the values given above for $D_0$(CN) and $D_0(N_2)$ leads to 9.77 e. v. for $D_0$(CO), and hence to 140 kcal/mole for the heat of sublimation of graphite and 86 kcal./mole for the dissociation energy of cyanogen.

The methods that we have used in reaching these values involve some novel ideas—the concept of valence-state energy and the postulate that the valence-state energy is usually nearly the same for an atom in different molecules, the use of linear extrapolation of vibrational levels to obtain the dissociation energies of molecules into atoms in their valence states, the use of conjugation energy for expressing a difference in dissociation energies—and their reliability and accuracy have not been extensively tested. From internal evidence it seems likely, however, that these methods when cautiously applied yield energy values that involve no larger errors than a few tenths of an electron-volt, and that they may be used with considerable confidence, as in the present paper, to select the correct ones from among alternative widely differing possible values of dissociation energies of molecules.

---

[*] Contribution No. 1297.

[1] Herzberg, G., *Chem. Rev.*, **20,** 145–167 (1937); *J. Chem. Phys.*, **10,** 306–307 (1942).

[2] Hagstrum, H. D., and Tate, J. T., *Phys. Rev.*, **59,** 354–370 (1941); Hagstrum, H. D., *Ibid.*, **72,** 947–963 (1947).

[3] Coster, D., and Brons, F., *Physica*, **1,** 155–160 (1934).

[4] Gaydon, A. G., and Penney, W. G., *Proc. Roy. Soc. London*, **183A,** 374–388 (1944).

[5] Brewer, L., Gilles, P. W., and Jenkins, F. A., *J. Chem. Phys.*, **16,** 797–807 (1948).

[6] Long, L. H., *Ibid.*, **16,** 1087 (1948).

[7] Voge, H. H., *Ibid.*, **16,** 984–986 (1948).

[8] Brewer, L., *Ibid.*, **16,** 1165–1166 (1948).

[9] Pauling, L., *Z. Naturforschung*, **3A,** 438–447 (1948).

[10] Pauling, L., PROC. NATL. ACAD. SCI., **35,** 229–232 (1949).

[11] Mulliken, R. S., *Phys. Rev.*, **46,** 144–146 (1934).

[12] Wulf, O. R., *Ibid.*, **46,** 316 (1934).

[13] Hagstrum, H. D., *J. Chem. Phys.*, **16,** 848–849 (1948).

[14] Gaydon, A. G., *Dissociation Energies*, Chapman and Hall, Ltd., London, 1947.

[15] *Ibid.*, p. 182.

[16] Kistiakowsky, G. B., and Gershinowitz, H., *J. Chem. Phys.*, **1,** 432–439, 885 (1933).

[17] White, J. U., *Ibid.*, **8,** 459–465 (1940).

[18] Pauling, L., and Sherman, J., *Ibid.*, **1,** 679–686 (1933).

# AGGREGATION PHENOMENA IN POLYVINYL ALCOHOL-ACETATE SOLUTIONS*

By S. N. Timasheff, M. Bier and F. F. Nord

Department of Organic Chemistry,† Fordham University, New York 58, N. Y.

Communicated by F. G. Keyes, May 7, 1949

In a series of earlier studies it was found that freezing causes a change in the particle size of various enzymes, proteins and other colloids in aqueous solution by inducing aggregation or disaggregation.[1] Applying the property of light scattering to particle weight determination made it possible to extend investigations on the effect of temperature to the range of 20 to 50°C. It was found that at these temperatures recrystallized egg albumin is subject to an aggregation reaction within the alleged pH range of stability.[2] The light scattering method has been shown to be particularly well suited for the study of such phenomena, and was, therefore, further adopted for the investigation of changes in the state of aggregation of polyvinyl alcohol-acetate copolymers (PVA-A) in aqueous solutions.

Through partial hydrolysis of polyvinyl acetate, it is possible to obtain PVA-A copolymers of varying acetyl contents. The properties of these copolymers, notably their solubility, depend largely on the residual acetyl groups. Samples having an acetyl content of approximately 20–25% are soluble in cold water but precipitate out on heating. The diagram of solubility and of phase separation is, therefore, inverted and, contrary to the usual case, the system possesses a critical temperature below which there is total miscibility of water and PVA-A at all concentrations, and above which a separation of phases takes place. It seemed, therefore, of interest to investigate the phenomena preceding phase separation and to interpret these in terms of the thermodynamics of high polymer solutions.

For the calculation of the particle weight, $M$, from the turbidity, $\tau$, Debye's equation[3]

$$\frac{Hc}{\tau} = \frac{1}{M} + 2Bc \qquad (1)$$

was applied, where $c$ is the concentration, and $H$ is a function of the wavelength and the refractive indices of the solvent and solution. The constant $B$ depends upon the heat of mixing, and, according to Huggins,[4] is:

$$B = \frac{1}{V_1 d_2^2}\left(\tfrac{1}{2} - \mu\right) \qquad (2)$$

where:

$V_1$ = molar volume of solvent
$d_2$ = density of solute
$\mu$ = a function of the heat of mixing

Equation 1 is limited to a system of independent, dielectric, spherical particles of small diameter as compared with the wave-length of the light. For larger particles a correction factor must be calculated from the measured dissymmetry of the angular distribution of the scattered light,[3, 5] which also permits the calculation of the particle size.

*Experimental.*—For the determination of the turbidity and the dissymmetry of the scattered light, the previously applied Tyndallometer was used.[2] The specific refractive index increment of the solutions was determined with a Differential Refractometer of the type described by P. P. Debye,[6] using a cell of improved design.

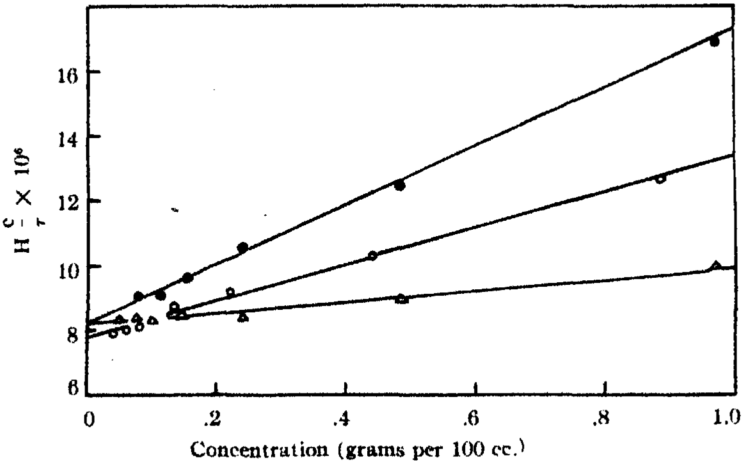


FIGURE 1

Values of $Hc/\tau$ of polyvinyl alcohol-acetate copolymer in aqueous solution vs. solute concentration at various temperatures. ●, 30°C.; ○, 40°C.; △, 50°C.

The PVA-A (E. I. du Pont de Nemours and Co., medium viscosity, Type A, Grade 52-22, lot JB 276) was brought into solution and dialyzed for 24 hrs. against running distilled water. In order to eliminate a small fraction, insoluble at the working temperatures, the solution was brought to 55°C. for 24 hrs., and then filtered at the same temperature through a Seitz sterilizing filter. The resulting 1% solution was subsequently stored at 4°C. until use.

Prior to each series of measurements the required quantity of the solution was kept for 24 hrs. in a thermostat at the desired temperature, and subsequently pressure-filtered at the same temperature through a specially designed sintered glass filter of fine porosity. The solution was then transferred to a series of semioctagonal cells used in the Tyndallometer and diluted to the working concentrations with filtered distilled water of the same temperature. The cells were introduced into the same thermostat and periodic measurements were taken. At the same time the index of refraction increment and the concentration were determined.

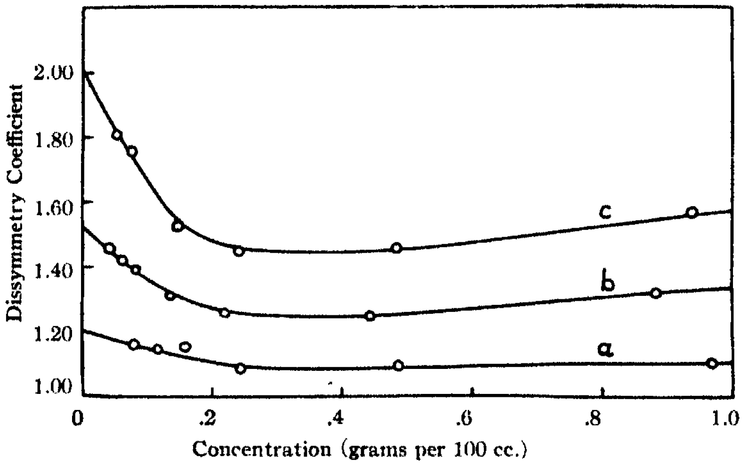


FIGURE 2

Dissymmetry coefficients (defined as ratio of the intensity of the light scattered at angles of 45° and 135°) of polyvinyl alcohol-acetate copolymer in aqueous solution vs. concentration of the solute at various temperatures. *a*, 30°C.; *b*, 40°C.; *c*, 50°C.

TABLE I

| TEMP., °C. | $1/M \times 10^6$, UNCORRECTED | DISSYMMETRY COEFFICIENT, $i_{45}/i_{135}$ | CORRECTION FACTOR | PARTICLE WEIGHT $M$, CORRECTED |
|---|---|---|---|---|
| 30 | 8.24 | 1.20 | 1.14 | 138,000 |
| 40 | 7.84 | 1.52 | 1.38 | 176,000 |
| 50 | 8.20 | 2.01 | 1.82 | 222,000 |

*Results and Discussion.*—Uncorrected turbidity-concentration curves for the PVA-A solutions at temperatures of 30°, 40° and 50°C. are shown in figure 1, the concentrations ranging from 0.04 to 1%. Corresponding dependence of the dissymmetry of the scattered light on the concentration is presented in figure 2. The reported values are average values of measurements taken over a period of 24 hrs., during which time no change in values exceeding the experimental error was observed.

From the presented graphs it is evident that a far-reaching aggregation has already occurred at temperatures well below the critical point of phase separation. Both the average particle weight and the particle size have increased with rising temperature. The corresponding values are recorded in table 1.

Accompanying the increase in particle weight, the slopes of the $Hc/\tau$ *vs.* concentration graphs have shown a notable decrease with rise in temperature. In table 2 are recorded the values of the slope B (eq. 1), uncorrected and corrected for the dissymmetry, and the corresponding values of $\mu$ (eq. 2) at the three temperatures. With rising temperature there is a noticeable increase in the factor $\mu$, which, according to the thermodynamic treatment of polymer solutions by Flory[7] and Huggins,[4] indicates the approach of the critical point of phase separation. At this point $\mu$ should assume a value of 0.5 or slightly higher, the slope B (eq. 1) becoming zero.

TABLE 2

| TEMP., °C. | $B \times 10^4$, UNCORRECTED | CORRECTION FACTOR | $B \times 10^4$, CORRECTED | $\mu$ |
|---|---|---|---|---|
| 30 | 4.48 | 1.14 | 5.12 | 0.485 |
| 40 | 2.77 | 1.38 | 3.83 | 0.489 |
| 50 | 0.77 | 1.82 | 1.40 | 0.496 |

NOTE: Specific gravity of PVA-A = 1.259.

This thermodynamic treatment should not be applied, however, to aqueous solutions without reserve. Nevertheless, $\mu$ can be considered an empirical constant characteristic of a given solute-solvent system,[8] and may be expressed in the form:

$$\mu = \beta + \frac{\alpha}{RT} \tag{3}$$

where $\alpha$ and $\beta$ are empirical constants, the heat of mixing being the most important factor of $\alpha$. As, in solutions of PVA-A, the factor $\mu$ increases with a rise in temperature, it is evident from the derivative of equation 3 with respect to $T$, that in this case $\alpha$ must be negative, i.e., the heat of mixing is negative.

*Summary.*—Applying the light scattering method, the dependence on temperature of particle weights of polyvinyl alcohol-acetate copolymers in aqueous solution was determined. An increase in temperature causes a far reaching aggregation, accompanied by an increase in the value of the factor $\mu$. The aggregation, dependent upon a negative heat of mixing of the polymer with water, can be treated in terms of the thermodynamics of phase separation.

---

* This work has been carried out under the auspices of the Office of Naval Research.

† Communication No. 178.

[1] Nord, F. F., *Science*, **75**, 54 (1932), *Ergebn. d. Enzymforschg.*, **2**, 23 (1933); Dammann, E., Lange, F. E. M., Bredig, M. A., and Nord, F. F., *Biochem. Z.*, **228**, 421 (1936);

Endoh, C., Lange, F. E. M., and Nord, F. F., *Ber.*, **68**, 2004 (1935); Leichter, H., and Nord, F. F., *Biochem. Z.*, **295**, 226 (1938); Holzapfel, L., and Nord, F. F., *Ber.*, **71**, 1217 (1938); *Biodynamica*, **3**, 1 (1940); Holzapfel, L., *Koll. Z.*, **85**, 272 (1938); Rampino, L. D., and Nord, F. F., *J. Am. Chem. Soc.*, **63**, 2745 (1941).

[2] Bier, M., and Nord, F. F., PROC. NATL. ACAD. SCI., **35**, 17 (1949).

[3] Debye, P., *J. Phys. and Colloid Chem.*, **51**, 18 (1947).

[4] Huggins, M. L., *Ann. N. Y. Acad. Sci.*, **43**, 1 (1942); *Ind. Eng. Chem.*, **35**, 216 (1943).

[5] Zimm, B. H., Stein, R. S., and Doty, P. M., *Polymer Bull.*, **1**, 90 (1945); Stein, R. S., and Doty, P. M., *J. Am. Chem. Soc.*, **68**, 159 (1946).

[6] Debye, P. P., *J. Applied Phys.*, **17**, 392 (1946).

[7] Flory, P. J., *J. Chem. Phys.*, **10**, 51 (1942); **13**, 453 (1945).

[8] Huggins, M. L., *Ann. N. Y. Acad. Sci.*, **44**, 431 (1943).

# DIFFUSION EQUATION AND STOCHASTIC PROCESSES

BY S. BOCHNER

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY

Communicated by Marston Morse, May 7, 1949

Take a point set $S$ with a Lebesgue measure and consider on it a stochastic process with a density function $f(s, x; t, y)$, $x, y \in S$, $0 < s < t < \infty$. It satisfies the functional equation

$$f(s, x; t, y) = \int_S f(s, x; u, z)\, f(u, z; t, y) dv_z$$

$0 < s < u < t$, with the accompanying initial condition

$$f(s, x; t, y) \to \delta(x, y) \text{ as } t - s \to 0$$

where $\delta(x, y)$ is the "Dirac function." By a general argument it can be shown that this always implies a pair of diffusion equations

$$\begin{aligned} \frac{\partial}{\partial s} f(s, x; t, y) &= -A_s(x) f(s, x; t, y) \\ \frac{\partial}{\partial t} f(s, x; t, y) &= A_t^*(y) f(s, x; t, y) \end{aligned} \qquad (1)$$

(one "backward" and one "forward" equation) in which, for each $u$, the symbol $A_u(z)$ denotes a suitable linear operator acting on the point $z$, and the symbol $A_u^*(z)$ denotes its "adjoint" operator. If $f(s, x; t, y)$ is stationary, that is if there exists a function $f(u; x\ y)$ such that $f(s\ x; t\ y) = f(t - s; x\ y)$ then there exists one operator $A(z)$ so that $A_u(z) = A(z)$ and

$$\frac{\partial}{\partial u} f(u; x\ y) = A(x)\, f(u; x, y). \qquad (2)$$

All this does not yet exhaust the description of a stochastic process. What has to be added is the very important property

$$f(s, x; t, y) \geq 0 \tag{3}$$

which has an initial condition of its own namely

$$\int_S f(s, x; t, y) dv_y = 1. \tag{4}$$

Condition (4) may be very troublesome, either to enforce or to verify, but there is no intuitive difficulty in describing it in terms of the generating operator $A_u(z)$. Thus, for a stationary process $A(z)$, if $A(z)$ is a self-adjoint operator on $L_2$ it states that $A$ shall have the eigenvalue $\lambda_0 = 0$ with the eigenfunction $\varphi_0(x) = 1$ or approximately so. However the intuitive reduction of the "global" property (3) to a structural property of the generating operator is in the general case much harder to accomplish.

Actually, in the classical set-up with which the theory of stochastic processes started, such a reduction does exist, and very sharply so. In the classical stationary set-up the operator $A$ is the negative Laplace-Beltrami elliptic differential operator of the second order, $A = -\Delta$, on a differentiable manifold, and property (3) follows then, sufficiently, from the fact, that at a point of minimum for a twice differentiable function we have $\Delta g \geq 0$. So this is quite satisfactory as far as it goes. But the Laplace operator leads only to a Gaussian stochastic process, or approximately so, and for all other processes the problem remains.

Now, we claim that for a certain type of other processes an answer is still possible. Take a numerical function $\Phi(\lambda)$ in $0 \leq \lambda < \infty$ and assume that $\exp\{-u\Phi(\lambda)\}$ is completely monotone, that is

$$e^{-u\Phi(\lambda)} = \int_0^\infty e^{-\lambda\rho}\, d\gamma_u(\rho), \qquad d\gamma(\rho) \geq 0, \tag{5}$$

and if $B$ is any stationary process, say $B = -\Delta$, then we form the operational function $A = \Phi(B)$. We claim that it again forms a stationary process; in fact their densities satisfy the relation

$$f_\Phi(u; x, y) = \int_0^\infty f(\rho; x, y) d\gamma_u(\rho), \tag{6}$$

so that (3) implies $f_\Phi \geq 0$, as claimed.

In particular for $\Phi(\lambda) = \lambda^p$, $0 < p < 1$, we obtain the "stable" processes of P. Levy, and we have just seen that Levy's construction may be superimposed on any stationary process originally given and not only a Gaussian one. Thus, any diffusion equation (2) implies the associated equations

$$\frac{\partial}{\partial u} f = A^p f, \qquad 0 < p < 1, \tag{7}$$

all of diffusion type, and many others. We yet mention, for any $c > 0$, the associated equation

$$\frac{\partial}{\partial u} f = \{(A + c)^{1/2} - c^{1/2}\} f \tag{8}$$

which is the "relativistic" counterpart to (2), this counterpart being always a stochastic process again. However, their Brownian motions may be very different from each other. Thus, for $A = -\Delta$ almost all paths are infinite, but for $\Phi(-\Delta)$ almost all paths will remain infinite if and only if $\int_1^\infty \Phi(\lambda^2)\lambda^{-2}d\lambda = \infty$; otherwise, and this then includes the equation (7), for $p < 1/2$ they are almost all finite. Whether this might have a physical interpretation, is not known to us.

We can also make a statement on non-stationary cases. Take a function in two variables $\Phi(u, \lambda)$, $0 < u < \infty$, $0 \leq \lambda < \infty$, and assume that for every fixed $u$ it is completely continuous as before. Take a fixed self-adjoint operator $B$ which produces a stationary process, say $B = -\Delta$, and set up the variable operator

$$A_u = \Phi(u, B). \tag{9}$$

We claim that this will produce a non-stationary process in (1). It is no longer possible to substantiate this by a formula as direct as (6) is, but if we employ the spectral resolution of $B$, a formula can again be written down. Assume first that $B$ has a pure point spectrum in which case we have

$$f(u; x, y) = \sum_n e^{-u\lambda_n}\varphi_n(x)\varphi_n(y). \tag{10}$$

For the stochastic density of $A_u$ we then have

$$f(s, x; t, y) = \sum_n \exp\left(-\int_s^t \Phi(u, \lambda_n)du\right)\varphi_n(x)\varphi_n(y), \tag{11}$$

and it can be shown that the positiveness of (10) induces that of (11). More generally, for a non-discrete spectrum if we have a Carleman representation

$$f(u; x, y) = \int_0^\infty e^{-u\lambda}\, d_\lambda E(\lambda; x, y),$$

then this implies

$$f(s, x; t, y) = \int_0^\infty \exp\left(-\int_s^t \Phi(u, \lambda)du\right)d_\lambda E(\lambda; x, y).$$

Finally we note that if we start with a family of self adjoint operators $A_u$, $0 < u < \infty$, instead of investigating directly the existence of a function $\Phi(u, \lambda)$ and an operator $B$ for which (9) holds, it suffices to verify that they are mutually commutative, $A_uA_v = A_vA_u$, and that, for each $u$, $A_u$ generates a stationary process; because this then implies the representation (9) automatically.

A detailed paper with full references will appear at a later date.

# ON EPSTEIN'S ZETA FUNCTION (I)

BY S. CHOWLA AND A. SELBERG

INSTITUTE FOR ADVANCED STUDY, PRINCETON, N. J.

Communicated by H. Weyl, May 18, 1949

1. This paper contains a short account of results whose detailed proofs will be published later.

We define the function $Z(s)$ by

$$Z(s) = \sum{}'(am^2 + bmn + cn^2)^{-s} \tag{1}$$

where $s = \sigma + it$($\sigma$ and $t$, real), $\sigma > 1$, and the summation is for all integers $m$, $n$ (each going from $-\infty$ to $+\infty$), while the dash indicates that $m = n = 0$ is excluded from the summation; further $a$ and $c$ are positive numbers while $b$ is real and subject to $4ac - b^2 = \Delta > 0$.

It is well known that the function $Z(s)$, defined for $\sigma > 1$ by (1), can be continued analytically over the whole $s$-plane, and satisfies a functional equation similar to the one satisfied by the Riemann Zeta Function. The function $Z(s)$, thus defined, is a meromorphic function with a simple pole at $s = 1$.

Deuring (*Math. Ztschr.*, **37**, 403–413 (1933)) obtained an important formula for $Z(s)$. Deuring's work led Heilbronn (*Quart. J. Maths., Oxford*, **5**, 150 (1934)) to the proof of the following famous conjecture of Gauss on the class-number of binary quadratic forms with a negative fundamental discriminant: let $h(-\Delta)$ denote the number of classes of binary quadratic forms of negative fundamental discriminant $-\Delta = b^2 - 4ac$, then

$$h(-\Delta) \to \infty \qquad \text{as} \qquad \Delta \to \infty \tag{2}$$

Again using the ideas of Heilbronn and Deuring, Siegel proved that

$$h(-\Delta) > \Delta^{1/2 - \epsilon} \; [\Delta > \Delta_0(\epsilon)] \tag{3}$$

which is a great advance on (2).

Our starting point is the formula:

$$Z(s) = 2\zeta(2s)a^{-s} + \frac{2^{2s}a^{s-1}\sqrt{\pi}}{\Gamma(s)\,\Delta^{s-1/2}}\zeta(2s-1)\Gamma(s - 1/2) + Q(s) \tag{4}$$

where

$$Q(s) = \frac{\pi^s \cdot 2^{s+3/2}}{a^{1/2}\Gamma(s)\,\Delta^{s/2-1/4}} \sum_{n=1}^{\infty} n^{s-1/2}\,\sigma_{1-2s}(n)\cos\left(\frac{n\pi b}{a}\right)\int_0^\infty \phi^{s-3/2} \exp\left\{-\frac{\pi n\Delta^{1/2}}{2a}(\phi + \phi^{-1})\right\} d\phi \tag{4}$$

Here $\sigma_k(n)$ denotes the sum of the $k$th powers of the divisors of $n$, and $\zeta(s)$ is Riemann's Zeta Function. The series for $Q(s)$ is highly convergent. Taking a crude estimate of the series for $Q(s)$ we obtain the formula of Deuring referred to above.

2. The formula (4) can be applied to the proof of the positiveness of certain Dirichlet $L$-functions at $s = {}^1/_2$. In fact we define for $s > 0$,

$$L_p(s) = \sum_1^\infty \left(\frac{n}{p}\right) n^{-s}$$

where $(n/p)$ is Legendre's symbol defined as follows:

> If $n \not\equiv 0 \pmod{p}$, then $(n/p) = +1$ if the congruence $x^2 \equiv n \pmod{p}$ is soluble; $(n/p) = -1$ if the congruence $x^2 \equiv n \pmod{p}$ is insoluble
>
> If $n \equiv 0(p)$, then

$$(n/p) = 0$$

The positiveness of $L_p(s)$ for $0 < s \leqq 1$ was proved by S. Chowla (*Acta Arithmetica*, Band **1**, 114 (1935)) in a large number of special cases, e.g., for $p = 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 47, 53, 59, 61, 71, 73, 79, 83, 89, 97$.

But no information was obtained in the cases $p = 43, 67, 163$ (here the class number $h(-p)$ is small). Heilbronn (*Acta Arithmetica*, Band **2**, 212 (1937) proved that there are infinitely many primes $p$ for which the method of Chowla gives no information. Curiously enough, the present method is more successful with precisely those cases like $p = 43, 67, 163$ (class number $h(-p) = 1$) where the previous method failed. In these three cases we obtain $L_p({}^1/_2) > 0$ (Rosser has recently, in an unpublished paper, settled the cases $p = 43$ and $p = 67$ by an entirely different method). That $L_p({}^1/_2) > 0$ in these cases, is not surprising, for if there is a prime $p$ such that $L_p({}^1/_2) < 0$ then the extended Riemann hypothesis is false! These results are deduced from the following

THEOREM: *If $p$ is an odd prime $>7$ and if $h(-p) = 1$, then* $(c = \pi/2)$

$$\zeta({}^1/_2)L_p({}^1/_2) = \gamma + \log\left(\frac{\sqrt{p}}{8\pi}\right) + \frac{8\theta . e^{-c\sqrt{p}}}{\pi\sqrt{p}(1 - e^{-c\sqrt{p}})} \qquad (5)$$

*where $\gamma$ is Euler's constant and $\theta$ is a real number such that* $|\theta| < 1$. *Remark that we can also show the positivity of $L_p(\sigma)$ on the whole stretch ${}^1/_2 \leq \sigma \leq 1$ by the same method, in the three cases* $p = 43, 67, 163$. *This can be done with a little more computation.*

3. It is well known that we have

$$h(-d) = 1 \qquad (6)$$

in the nine cases $d = 3, 4, 7, 8, 11, 19, 43, 67, 163$. Heilbronn and Linfoot have proved that (6) has at most 10 solutions; further, that if (6) has a tenth solution then $d$ must be very large indeed; in fact $d > 5.10^9$ (Lehmer).

It follows from (5) that if (6) has a tenth solution $d = d_0$, then

$$L_d(^1/_2) < 0 \qquad [d = d_0]$$

It is known that $d_0$ is necessarily a prime, and is in fact, $\equiv 3 \pmod 8$.

4. We apply (4) to a classical problem of the theory of elliptic functions. Write, as usual,

$$K = \int_0^{\pi/2} \frac{d\phi}{\sqrt{1 - k^2 \sin^2 \phi}} \qquad (0 < k < 1)$$

$$K' = \int_0^{\pi/2} \frac{d\phi}{\sqrt{1 - k'^2 \sin^2 \phi}} \qquad (k^2 + k'^2 = 1)$$

It has long been known that $K$ can be calculated in finite terms whenever $iK'/K$ is a number belonging to any of the imaginary quadratic fields $k(\sqrt{-1})$, $k(\sqrt{-2})$, $k(\sqrt{-3})$. This is deduced from the fact that $K$ can be calculated in finite terms when

$$K'/K = \sqrt{n} \qquad (n = 1, 2, 3)$$

Thus when $n = 1$ i.e., $k = \dfrac{1}{\sqrt{2}}$, we have

$$K = \frac{\Gamma^2(^1/_4)}{4\sqrt{\pi}}$$

and there are similar results obtained (each by a different method!) for the cases $n = 2, 3$. We prove that $K$ can be calculated in finite terms whenever $iK'/K$ is a number in an imaginary quadratic field. More precisely, our result is as follows: let $d$ be a negative integer $\equiv 0$ or $1 \pmod 4$ and so that $d$ or $d/4$ is a square-free integer. Further let $h$ denote the class-number $h(d)$ and

$$w = 6, 4, 2 \qquad \text{according as } d = -3,\ d = -4,\ d < -4.$$

Finally $(d/m)$ denotes the Kronecker symbol. Then,

THEOREM: *Let $iK'/K$ be a number from the field $k(\sqrt{d})$. Then we have*

$$K = \lambda\sqrt{\pi} \left\{ \mathop{\pi}\limits_{m=1}^{|d|} \Gamma\left(\frac{m}{|d|}\right)^{(d/m)} \right\}^{w/4h} \tag{7}$$

*where $\lambda$ is an algebraic number.*

A special case of (7) is the following:

THEOREM: *If $K'/K = \sqrt{p}$ and if $h(-p) = 1$, then*

$$\frac{2K}{\pi} = \frac{2^{1/3}(kk')^{-1/6}}{\sqrt{2\pi p}} \left\{ \frac{\pi\, \Gamma\left(\frac{\alpha}{p}\right)}{\pi\, \Gamma\left(\frac{\beta}{p}\right)} \right\}^{w/4} \qquad (8)$$

*where (p is a prime) and* $w = 6$ *if* $p = 3$, $w = 2$ *if* $p > 3$; $\alpha$ *runs through the* $\frac{p-1}{2}$ *quadratic residues of p that lie between* 0 *and p, while* $\beta$ *runs through the remaining* $\frac{p-1}{2}$ *numbers between* 0 *and p.*

Specializing again to the case $p = 7$ we obtain in the usual notation for hypergeometric series:

$$F(^1/_4, {}^1/_4, 1; {}^1/_{64}) = \sqrt{\frac{2}{7\pi}} \left\{ \frac{\Gamma(^1/_7)\Gamma(^2/_7)\Gamma(^4/_7)}{\Gamma(^3/_7)\Gamma(^5/_7)\Gamma(^6/_7)} \right\}^{1/2}$$

5. Let $G_d(s)$ denote the analytical continuation of the function defined for $\sigma > {}^3/_2$ by the series

$$\sum{}'(x^2 + y^2 + dz^2)^{-s}$$

From a formula similar to (4) it is deduced that

THEOREM: *There exists a real number* $\theta_d$ *such that*

$$G_d(\theta_d) = 0 \qquad [d > d_0]$$

*where* $\theta_d \to 0$ *as* $d \to \infty$, *but* $\theta_d \neq 0$.

# ON A NEW METHOD IN ELEMENTARY NUMBER THEORY WHICH LEADS TO AN ELEMENTARY PROOF OF THE PRIME NUMBER THEOREM

BY P. ERDÖS

DEPARTMENT OF MATHEMATICS, SYRACUSE UNIVERSITY

Communicated by P. A. Smith, April 16, 1949

1. *Introduction.*—In the course of several important researches in elementary number theory A. Selberg[1] proved some months ago the following asymptotic formula:

$$\sum_{p \leqq x} (\log p)^2 + \sum_{pq \leqq x} \log p \log q = 2x \log x + O(x), \qquad (1)$$

where $p$ and $q$ run over the primes. This is of course an immediate consequence of the prime number theorem. The point is that Selberg's ingenious proof of (1) is completely elementary. Thus (1) can be used as a starting point for elementary proofs of various theorems in analytical number theory, which previously seemed inaccessible by elementary methods.

Using (1) I proved that $p_{n+1}/p_n \to 1$ as $n \to \infty$. In fact, I proved the following slightly stronger result: To every $c$ there exists a positive $\delta(c)$, so that for $x$ sufficiently large we have

$$\pi[x(1 + c)] - \pi(x) > \delta(c)x/\log x \tag{2}$$

where $\pi(x)$ is the number at primes not exceeding $x$.

I communicated this proof of (2) to Selberg, who, two days later, using (1), (2) and the ideas of the proof of (2), deduced the prime number theorem $\lim\limits_{x \to \infty} \frac{\pi(x)\log x}{x} = 1$ or, equivalently[2]

$$\lim_{x \to \infty} \frac{\vartheta(x)}{x} = 1, \text{ where } \vartheta(x) = \sum_{p \leq x} \log p. \tag{3}$$

In a few more days, Selberg simplified my proof of (2), and later we jointly simplified the proof of the prime number theorem. The new proof no longer required (2), but used the same ideas as in the proof of (2) and (3). I was also able to prove the prime number theorem for arithmetic progressions. My proof of the latter was helped by discussions with Selberg and it utilizes ideas of Selberg's previous elementary proof of Dirichlet's theorem,[3] according to which every arithmetic progression whose first term and difference are relatively prime contains infinitely many primes. This proof will be given in a separate paper.

Selberg has now a more direct proof of (3), which is not yet published. It is possible, therefore, that the present method may prove to be only of historical interest.

I now proceed to give the proofs as they occurred in chronological order. (It should be remarked that we never utilize the full strength of (1), indeed an error term $o(x \log x)$ is all that is used in the following proofs.)

We introduce the following notation:

$$A = \limsup_{x \to \infty} \frac{\vartheta(x)}{x}, \quad a = \liminf_{x \to \infty} \frac{\vartheta(x)}{x}.$$

First, we state a few elementary facts about primes which will be used subsequently. Of these, I, II and IV are well known in elementary prime number theory, while III is shown to be a simple consequence of (1).

I. $a > 0$.

II. $\sum\limits_{p \leq x} \frac{\log p}{p} = [1 + o(1)]\log x.$

III. Let $x_2 > x_1$. Then

$$\vartheta(x_2) - \vartheta(x_1) < 2(x_2 - x_1) + o(x_2).$$

Thus, in particular, if $x_1 = 0$, we obtain $A \leqq 2$.

Put in (1) $x = x_2$ and $x = x_1$ and subtract. Then we obtain

$$\sum_{x_1 < p \leqq x_2} (\log p)^2 \leqq 2x_2 \log x_2 - 2x_1 \log x_1 + o(x_2 \log x_2) \leqq 2(x_2 - x_1)\log x_2 + o(x_2 \log x_2). \quad (4)$$

We distinguish two cases: (A) $x_1 \geqq x_2/(\log x_2)^2$. Then clearly $\log x_1 = (1 + o(1))\log x_2$ and III follows from (4) on dividing both sides by $\log x_2$. (B) $x_1 < x_2/(\log x_2)^2$. Then we have by (A)

$$\vartheta(x_2) - \vartheta(x_1) < \vartheta(x_2) - \vartheta(x_2/(\log x_2)^2) + \frac{x_2}{(\log x_2)^2} \log x_2 < 2\left(x_2 - \frac{x_2}{(\log x_2)^2}\right) + o(x_2) = 2(x_2 - x_1) + o(x_2), \text{ q. e. d.}$$

IV. $A \leqq 1.5$. This is a consequence of the known result $\vartheta(x) \leqq 1.5x$.

2. *Proof of* (2).—It is equivalent to prove that to every positive $c$ there exists a positive $\delta(c)$ such that $\vartheta[(1 + c)x] - \vartheta(x) > \delta(c)x$ for $x$ sufficiently large.

Suppose this not true, then there exist positive constants $c'$ and corresponding arbitrarily large $x$ so that

$$\vartheta[x(1 + c')] - \vartheta(x) = o(x). \quad (5)$$

Put $C = \sup c'$. It easily follows from I and the finiteness of $A$ that $C < \infty$.

First we show that $C$ satisfies (5), in other words, that there are arbitrarily large values of $x$ for which

$$\vartheta[x(1 + C)] - \vartheta(x) = o(x). \quad (6)$$

Choose $c' > C - {}^1/_2\epsilon$ and let $x \to \infty$ through values satisfying (5). Then by III we have

$$\vartheta[x(1 + C)] - \vartheta(x) = \vartheta[x(1 + C)] - \vartheta[x(1 + c')] + \vartheta[x(1 + c')] - \vartheta(x) \leqq 2(C - c')x + o(x) < \epsilon x + o(x),$$

which (since $\epsilon$ can be chosen arbitrarily small) proves (6).

Now we shall show that (6) leads to a contradiction. From (1) we obtain by subtraction

$$\sum_{x < p \leqq x(1+C)} (\log p)^2 + \sum_{x < pq \leqq x(1+C)} \log p \log q = 2Cx \log x + o(x \log x).$$

From (6) we have for suitable $x$ since $\sum_{x < p \leqq x(1+C)} (\log p)^2 = o(x \log x)$

$$\sum_{p \leq x(1+X)} \log p\left(\vartheta\left[\frac{x}{p}(1+C)\right] - \vartheta\left(\frac{x}{p}\right)\right) = 2Cx\log x + o(x\log x) \quad (7)$$

Now we deduce the following fundamental lemma.

Lemma 1. *Let $x \to \infty$ through values satisfying (6), then for all primes $p \leq x(1 + C)$, except possibly for a set of primes for which*

$$\sum \frac{\log p}{p} = o(\log x) \quad (8)$$

*we have*

$$\vartheta\left[\frac{x}{p}(1+C)\right] - \vartheta\left(\frac{x}{p}\right) = 2C\frac{x}{p} + o\left(\frac{x}{p}\right). \quad (9)$$

Suppose the lemma is not true. Then there exist two positive constants $b_1$ and $b_2$ so that for arbitrarily large $x$ (satisfying (6)) we have for a set of primes satisfying $\sum_{p \leq x(1+C)} \frac{\log r}{r} \sim b_1 \log x$

$$\vartheta\left[\frac{x}{r}(1+C)\right] - \vartheta\left(\frac{x}{p}\right) < (2C - b_2)\frac{x}{r}. \quad (10)$$

But then from II, III and (10), since (9) holds at best for a set of primes satisfying $\sum \frac{\log r}{r} \sim (1 - b_1)\log x$ we have

$$\sum_{p \leq x(1+C)} \log p\left(\vartheta\left[\frac{x}{r}(1+C)\right] - \vartheta(x)\right) \leqq b_1(2C - b_2)x\log x +$$
$$2C(1 - b_1)x\log x + o(x\log x) = (2C - b_1b_2)x\log x + o(x\log x)$$

But this contradicts (7), hence the lemma is established.

The primes satisfying (9) we shall call good primes, the other primes we shall call bad primes (of course the goodness and badness of a prime depends on $x$).

We shall prove the existence of a sequence of good primes $p_1 < p_2 < \dots p_k$ satisfying the following conditions:

$$10p_1 < p_k < 100p_1,\ (1 + C)(1 + t)^2 p_i > p_{i+1} >$$
$$(1 + t)p_i,\ i = 1, 2, \dots, k - 1 \quad (11)$$

where $t$ is a small but fixed number (small compared to $C$). Since $(1 + t)^k < 100$ it is clear that $k < k_0$ with constant $k_0 = k_0(t)$.

Suppose we already established the existence of a sequence satisfying (11). Then we prove (2) as follows: Consider the two intervals

$$\left[\frac{x}{p_{i+1}}, \frac{x}{p_{i+1}}(1 + C)\right], \left[\frac{x}{p_i}, \frac{x}{p_i}(1 + C)\right] \tag{12}$$

If they overlap, then by (11)

$$\frac{x}{p_{i+1}}(1 + t) < \frac{x}{p_i} < \frac{x}{p_{i+1}}(1 + C)$$

Clearly

$$\vartheta\left(\frac{x}{p_i}\right) - \vartheta\left(\frac{x}{p_{i+1}}\right) = 2\left(\frac{x}{p_i} - \frac{x}{p_{i+1}}\right) + o\left(\frac{x}{p_i}\right) \tag{13}$$

since otherwise

$$\vartheta\left(\frac{x}{p_i}\right) - \vartheta\left(\frac{x}{p_{i+1}}\right) < (2 - c_1)\left(\frac{x}{p_i} - \frac{x}{p_{i+1}}\right)$$

with $c_1 > 0$ and we would have from (9)

$$\vartheta\left[\frac{x}{p_{i+1}}(1 + C)\right] - \vartheta\left(\frac{x}{p_i}\right) > (2 + c_2)\left[\frac{x}{p_{i+1}}(1 + C) - \frac{x}{p_i}\right]$$

which contradicts III. Adding (13) and (9) with $p = p_i$ we obtain

$$\vartheta\left[\frac{x}{p_i}(1 + C)\right] - \vartheta\left(\frac{x}{p_{i+1}}\right) = 2\left[\frac{x}{p_i}(1 + C)\right] - \frac{x}{p_{i+1}} + o\left(\frac{x}{p_i}\right) \tag{14}$$

If the intervals (12) do not overlap we obtain by a simple calculation (using (9) and the fact that $t$ is small)

$$\vartheta\left[\frac{x}{p_i}(1 + C)\right] - \vartheta\left(\frac{x}{p_{i+1}}\right) > 1.9\left[\frac{x(1 + C)}{p_i} - \frac{x}{p_{i+1}}\right] \tag{15}$$

Adding all the equations (14) and (15) (for $i = 1, 2, \ldots, k$) we clearly obtain

$$\vartheta\left[\frac{x}{p_1}(1 + C)\right] - \vartheta\left(\frac{x}{p_k}\right) > 1.9\left[\frac{x}{p_1}(1 + C) - \frac{x}{p_k}\right] \tag{16}$$

Since $p_k > 10p_1$ we obtain from (16)

$$\vartheta\left[\frac{x}{p_1}(1 + C)\right] > 1.6\frac{x}{p_1}(1 + C). \tag{17}$$

But (17) contradicts IV.

Thus to complete the proof of (2) it will suffice to show the existence of a sequence of good primes satisfying (11).

Consider the intervals

$$I_r = (B^{2r}, B^{2r+1}),\ r = 0, 1, \ldots, \left[\frac{\log x}{2 \log B}\right] - 1,$$

where $B$ is a fixed, sufficiently large number. Clearly all the intervals $I_r$ lie in the interval $(0, x)$. First we show that with the exception of $o(\log x)$ $r$'s the interval $I_r$ contains good primes. From I and IV it easily follows that for sufficiently large $B$ we have (since $\vartheta(Bx) - \vartheta(x) > cx$)

$$\sum_{p \text{ in } I_r} \frac{\log p}{p} > c_1 (c_1 > 0 \text{ independent of } r)$$

Thus if there were $c_2 \log x$ with $c_2 > 0$ of the $I_r$'s without good primes, we would have

$$\sum_{p \text{ bad}} \frac{\log p}{p} > c_1 c_2 \log x$$

which contradicts (8).

Let now $p_1^{(r)}$ be the smallest good prime in $I_8$ (if it exists), and suppose that a sequence $p_1^{(r)}, p_2^{(r)}, \ldots, p_t^{(r)}$ satisfying (11) exists, but no $p_{t+1}^{(r)}$ satisfying (11) can be found. Thus, all the primes in

$$J_t^{(r)} = [p_t^{(r)} (1 + t), p_t^{(r)} (1 + t)^2 (1 + C)]$$

are bad. We have, by the definition of $C$,

$$\sum_{p \text{ in } J^{(r)}} \log p > \eta p_t^{(r)} (1 + t)^2 (1 + C), \ (\eta \text{ absolute constant}).$$

Thus

$$\sum_{p \text{ in } J^{(r)}} \frac{\log p}{p} > \eta \tag{18}$$

Clearly for $B > 100$ we have $p_i^{(r)}(1 + t)^2(1 + C) < B^{2r+2}$. Thus the intervals $J_{t_1}^{r_1}, J_{t_2}^{r_2}, \ldots$ do not overlap. Hence from (18), since the number of $r$'s with $p_1^{(r)}$ existing is $> \frac{\log x}{4 \log B}$,

$$\sum_{p \text{ bad}} \frac{\log p}{p} > \frac{\eta \log x}{4 \log B}$$

which contradicts (8) and establishes (2).

*3. Selberg's deduction of the prime number theorem from* (2).—Assume $a < A$. First we prove the following lemmas.

LEMMA 2. *$a + A = 2$.*

Choose $x \to \infty$ so that $\vartheta(x) = Ax + o(x)$. Then a simple computation (as in the proof of III) shows that

$$\sum_{p \leq x} (\log p)^2 = Ax \log x + o(x \log x).$$

Thus from (1)

$$\sum_{p \leqq x} (\log p)\vartheta\left(\frac{x}{r}\right) = (2 - A)x \log x + o(x \log x). \tag{19}$$

By the definition of $a$ and by II we obtain by a simple computation

$$\sum_{p \leqq x} (\log p)\vartheta\left(\frac{x}{r}\right) \geqq ax \sum_{p \leqq x} \frac{\log p}{p} + o(x \log x) = ax \log x + o(x \log x)$$

Thus from (19), $2 \geqq a + A$. We obtain $a + A \leqq 2$ similarly, by choosing $x$ so that $\vartheta(x) = ax + o(x)$. Thus lemma 2 is proved.

LEMMA 3. *Let $x \to \infty$ so that $\vartheta(x) = Ax + o(x)$. Then for any prime $p_i \leqq x$ except possible for a set of primes satisfying*

$$\sum \frac{\log p}{p} = o(\log x) \tag{20}$$

*we have*

$$\vartheta\left(\frac{x}{p_i}\right) = a\frac{x}{p_i} + o\left(\frac{x}{p_i}\right) \tag{21}$$

Suppose the lemma is false. Then as in the proof of lemma 1 there exist two positive constants $b_1$ and $b_2$ so that for arbitrarily large $x$, satisfying $\vartheta(x) = Ax + o(x)$, and for a set of primes satisfying $\sum \frac{\log p}{p} > b_1 \log x$, *we have*

$$\vartheta\left(\frac{x}{p}\right) > (a + b_2)\frac{x}{p} \tag{22}$$

But then we have from (22), lemma 2, (19) and II (as in the proof of lemma 1)

$$ax \log x + o(x \log x) = \sum_{p \leqq x} (\log p)\vartheta\left(\frac{x}{p}\right) > b_1(a + b_2)x \log x + (1 - b_1)ax$$

$$\log x + o(x \log x) = ax \log x + b_1b_2x \log x + o(x \log x),$$

an evident contradiction. This proves lemma 3.

LEMMA 4. *Let $p_1$ be the smallest prime satisfying (21). Then $p_1 < x^{\epsilon}$, and for all primes $p_j < x/p_1$ except possible for a set of primes satisfying*

$$\sum \frac{\log p}{p} = o(\log x) \tag{23}$$

*we have*

$$\vartheta\left(\frac{x}{p_1p_j}\right) = A\frac{x}{p_1p_j} + o\left(\frac{x}{p_1p_j}\right) \tag{24}$$

$p_1 < x^\epsilon$ follows immediately from (20) and II. The second part of lemma 4 follows by applying the argument of lemma 3 to $x/p_1$ instead of $x$ and interchanging $A$ and $a$.

Now the deduction of the prime number theorem. Let $p_t$ be any prime satisfying (21). Assume $\frac{x}{p_1p_j} < \frac{x}{p_t}$. Then (since $\vartheta(x)$ is non-decreasing) from (21) and (24)

$$a\frac{x}{p_t} + o\left(\frac{x}{p_t}\right) \leq A\frac{x}{p_1p_j} + o\left(\frac{x}{p_1p_j}\right)$$

or $p_j$ cannot lie in the interval

$$I_t = \left[\frac{p_t}{p_1}, \frac{p_t}{p_1}\left(\frac{A}{a} - \delta\right)\right],$$

where $\delta > 0$ is an arbitrary fixed number. Hence all primes in $I_t$ must be "bad," i.e., they do not satisfy (24). But it immediately follows from (2) that

$$\sum_{p \text{ in } I_t} \frac{\log p}{p} > \eta$$

To obtain a contradiction to (23) it suffices to construct $c \log x$ disjoint intervals $I_t$. This can be accomplished in the same way as in the end of the proof of (2) (where the disjoint intervals $J_t^{(r)}$ were constructed). This completes the first elementary proof of the prime number theorem.

4. *Sketch of Selberg's simplification of the proof of* (2).—If we can find two good primes satisfying

$$(1 + c)p_1 > p_2 > (1 + t)p_1, \quad c > \frac{C}{1 + t} \tag{25}$$

then (2) follows easily. The intervals $\left[\frac{x}{p_1}, \frac{x}{p_1}(1 + c)\right], \left[\frac{x}{p_2}, \frac{x}{p_2}(1 + c)\right]$, overlap. Thus (13), with $i = 1$, holds. But then exactly as in lemma 1 there exists a prime $p$ so that

$$\vartheta\left[\frac{x}{p_1p}(1 + c)\right] - \vartheta\left(\frac{x}{p_2p}\right) = o\left(\frac{x}{p_2p}\right).$$

But this is impossible (by the definition of $C$) since

$$\frac{x}{p_1p}(1 + c)/\frac{x}{p_2p} = \frac{p_2(1 + c)}{p_1} > 1 + C.$$

Thus we only have to show that good primes satisfying (24) exist, and this can be accomplished by using III (a contradiction with $\sum_{p \text{ good}} \frac{\log p}{p} = [1 + o(1)]\log x$ can be established similarly as in the previous proof).

*5. The joint simplified proof of the prime number theorem.*—

LEMMA 5. *Let $x_2 > x_1$ and $x_1 \to \infty$. Assume that $\vartheta(x_1) = Ax_1 + o(x_1)$ and $\vartheta(x_2) = ax_2 + o(x_2)$, or $\vartheta(x_1) = ax_1 + o(x_1)$ and $\vartheta(x_2) = Ax_2 + o(x_2)$. Then*

$$x_2/x_1 \leqq A/a + o(1).$$

Since $\vartheta(x)$ is non-decreasing we have in the first case

$$ax_2 + o(x_2) \leqq Ax_1 + o(x_1) \text{ or } x_2/x_1 \leqq A/a + o(1)$$

In the second case we have by III $\vartheta(x_2) - \vartheta(x_1) \leqq 2(x_2 - x_1) + o(x_2)$

$$ax_1 + 2(x_2 - x_1) \leqq Ax_2 + o(x_2) \text{ or } (2 - A)x_2 \leqq (2 - a)x_1 + o(x_2).$$

Hence by lemma 2, $ax_2 \leqq Ax_1 + o(x_2)$. Thus again $x_2/x_1 \leqq A/a + o(1)$. q.e.d.

Put $1 + D = \frac{A}{a} + \delta$ where $\delta$ is sufficiently small, and will be determined later. Next we prove the following result.

LEMMA 6.

$$\sum_{y \leqq p \leqq (1+D)y} \frac{\log p}{p} > \eta (\eta \text{ independent of } y).$$

First we show that

$$\sum_{y \leqq p \leqq (1+D)y} \log p > \eta(1 + D)y. \tag{26}$$

If (26) is false then for a suitable sequence of $y$'s we have $\vartheta[(1 + D)y] - \vartheta(y) = o(y)$. But then for these $y$'s

$$\frac{\vartheta[(1 + D)y]}{(1 + D)y} = \frac{\vartheta(y) + o(y)}{(1 + D)y} \leqq \frac{Ay + o(y)}{(1 + D)y} < a - c_1,$$

which contradicts the definition of $a$. Thus (26) holds and lemma 6 follows immediately.

Choose now $x$ so that $\vartheta(x) = Ax + o(x)$. Then by lemmas 3 and 4 we obtain ($p_1$, $p_i$ and $p_j$ having the same meaning as in lemmas 3 and 4)

$$\vartheta\left(\frac{x}{p_1p_j}\right) = A\frac{x}{p_1p_j} + o\left(\frac{x}{p_1p_j}\right), \quad \vartheta\left(\frac{x}{p_i}\right) = a\frac{x}{p_i} + o\left(\frac{x}{p_i}\right)$$

From lemma 5 we obtain that for any fixed $\epsilon$ and sufficiently large $x$ (satisfying $\vartheta(x) = Ax + o(x)$)

$$\text{either } \frac{x}{p_t} > \left(\frac{A}{a} - \epsilon\right)\frac{x}{p_1 p_j} \text{ or } \frac{x}{p_t} < \left(\frac{a}{A} + \epsilon\right)\frac{x}{p_1 p_j}$$

Hence $p_j$ cannot lie in the interval

$$I_t = \left[\left(\frac{a}{A} + \phantom{\epsilon}\right)\frac{p_t}{p_1}\left(\frac{A}{a} - \epsilon\right)\frac{p_t}{p_1}\right].$$

Now if $\delta$ is small enough then $1 + D \leqq \left(\frac{A}{a} - \epsilon\right)/\left(\frac{a}{A} + \epsilon\right)$. Hence by lemma 6

$$\sum_{p \text{ in } I_t} \frac{\log p}{p} > \eta$$

But by what has been said before all the primes in $I_t$ are bad (i.e., they do not satisfy (24)). Thus to arrive at a contradiction with (23) it will suffice as in the proof of (2) to construct $c \log x$ disjoint intervals $I_t$. This can be accomplished as in the proof of (2), which completes the proof of the prime number theorem.

6. Perhaps this last step can be carried out slightly more easily as follows: Put

$$S = \sum \frac{\log p_t}{p_t} \sum_{p \text{ in } I_t} \frac{\log p}{p} \tag{27}$$

where $p_t$ runs through the primes satisfying (21). As stated before all the primes in $I_t$ are bad (i.e., they do not satisfy (24)). Thus we have from (27)

$$S > \eta \sum \frac{\log p_t}{p_t} > \frac{\eta}{2} \log x \tag{28}$$

since by II and (20) $\Sigma \frac{\log p_t}{p_t} > {}^1/_2 \log x$ for large $x$.

On the other hand by interchanging the order of summation we obtain

$$S = \sum \frac{\log p}{p} \sum_{p \text{ in } J_p} \frac{\log p_t}{p_t}$$

where $p$ runs through all the primes of all the intervals $I_t$ (each $p$ is, of course, counted only once) and $p_t$ runs through the primes satisfying (21) of the interval

$$J_p = \left[pp_1\left(\frac{A}{a} - \epsilon\right)^{-1}, \quad pp_1\left(\frac{a}{A} + \epsilon\right)^{-1}\right].$$

We evidently have from $A < \infty$

$$\sum_{p_i \text{ in } J_p} \frac{\log p_i}{p_i} < c$$

Hence

$$\sum \frac{\log p}{p} > S/c$$

or from (27)

$$\sum \frac{\log p}{p} > \frac{\eta}{2c} \log x,$$

which contradicts (22) and completes the proof.

---

[1] Selberg's proof of (1) is not yet published.

[2] See, for example, Landau, E., *Handbuch der Lehre von der Verteilung der Primzahlen*, § 19, or Ingham, A. E., *The Distribution of Prime Numbers*, p. 13.

[3] An analogous result is used in Selberg's proof of Dirichlet's theorem.

[4] See, for example, Landau, E., op. cit., §§18 and 26, or Ingham, A. E., op. cit., pp. 14, 15 and 22.

---

# *ON THE STRUCTURE OF LOCALLY COMPACT GROUPS*

By Andrew M. Gleason

Society of Fellows, Harvard University

Communicated by Hassler Whitney, May 7, 1949

1. *Introduction*—Locally compact groups have attracted a great deal of study in the years since the introduction of invariant integration by Haar.[1] It has been shown that their structure is closely related to that of Lie groups in certain important cases (compact,[2] abelian[3] and solvable[4] groups), and it is widely conjectured that similar results are valid in general. We shall state here certain theorems which strengthen this conjecture and reduce its verification to the study of simple groups.

2. *The Extension Theorem for Lie Groups.*—THEOREM 1. *Let G be a topological group. Suppose that G has a closed normal subgroup N such that both N and G/N are Lie groups. Then G is itself a Lie group.*

In case $N$ is abelian, Kuranishi[5] has proved this theorem under the additional hypothesis that there is a local cross-section for the cosets of $N$; that is, a closed set having exactly one point in common with each coset of $N$ near the identity. The author has shown[6] that such a cross-section set always exists for abelian Lie groups. Hence our theorem is true for the special case that $N$ is abelian.

The remainder of the proof is straightforward. We may assume that $N$ is connected. Let $C$ be the centralizer of $N$. Then $C$ is a closed normal subgroup of $G$ and $G/C$ is a Lie group because it can be represented by automorphisms of $N$. It follows easily that $G/(C \cap N)$ is a Lie group. But $C \cap N$, being the center of $N$, is an abelian Lie group and we may apply the special case established above.

*3. Generalized Lie Groups.—Definition.* A topological group $G$ will be called a generalized Lie group if and only if the following condition is satisfied:

For every neighborhood $U$ of the identity, there is an open subgroup $G_1$ of $G$ and a compact normal subgroup $C$ of $G_1$ such that $C \subset U$ and $G_1/C$ is a Lie group.

A generalized Lie group is locally compact. The property of being a generalized Lie group is a local one; that is, if one of two locally isomorphic groups is a generalized Lie group, then so is the other.

THEOREM 2. *A locally compact group is a generalized Lie group if and only if it has an open subgroup which is the projective limit of Lie groups.*

It follows from this theorem that all compact groups and all locally compact abelian or totally disconnected groups are generalized Lie groups.

Any closed subgroup or quotient group of a generalized Lie group is again a generalized Lie group; in fact, the following analog of Cartan's theorem holds:

THEOREM 3. *Let $G$ be a locally compact group which admits a faithful representation in a generalized Lie group $H$. Then $G$ is a generalized Lie group.*

THEOREM 4. *Let $G$ be a topological group. Suppose $G$ has a closed normal subgroup $N$ such that both $N$ and $G/N$ are generalized Lie groups. Then $G$ is itself a generalized Lie group.*

It is an immediate corollary that every locally compact solvable group is a generalized Lie group. This result is slightly more inclusive than that of Malcev.[4]

*4. The Height of a Group.—Definition.* Let $G$ be a topological group. Consider increasing sequences $H_0 = \{e\} \subset H_1 \subset \ldots \subset H_k$ of connected closed subgroups of $G$ such that none of the factor spaces $H_i/H_{i-1}$ are compact. The **height** of $G$ is the least upper bound of the index $k$ taken over all such sequences. It is either a non-negative integer or $\infty$. We shall denote it by $h(G)$.

If $G_1$ is a closed subgroup of $G$, then $h(G_1) \leq h(G)$; moreover, if $N$ is a connected closed normal subgroup of $G$, then $h(N) + h(G/N) \leq h(G)$.

THEOREM 5. *The height of a locally compact group is finite.*

THEOREM 6. *Every locally compact group has a finite composition series such that all the quotient groups are either compact, abelian, totally disconnected, or simple.*

5. *Solvable and Semisimple Groups.*—THEOREM 7. *Every connected locally compact group has a largest normal solvable subgroup. The corresponding quotient group is semisimple; that is, it has no normal solvable subgroups except* $\{e\}$.

THEOREM 8. *Every connected locally compact semisimple group is the direct product of indecomposable groups. All of the factors except a finite number are compact simple Lie groups. The remaining factors are metrizable.*

6. *Conjecture.*—It seems natural to conjecture that every locally compact group is a generalized Lie group. To verify this conjecture, we may confine our attention to connected simple groups in view of Theorems 4 and 6. A connected simple group which is a generalized Lie group must be actually a Lie group. Hence our conjecture is equivalent to the assertion that every locally compact connected simple group is a Lie group.

If the conjecture can be established then Hilbert's fifth problem will be answered in the affirmative, for a connected generalized Lie group is a projective limit of Lie groups by Theorem 2 and a locally euclidean group which is a projective limit of Lie groups is known to be a Lie group.

[1] Haar, A., "Der Massbegriff in der Theorie der kontinuierlichen Gruppen," *Ann. Math.*, **34**, 147–169 (1933).

[2] Von Neumann, J., "Die Einführung analytischer Parameter in topologischen Gruppen," *Ibid.*, **34**, 170–190 (1933). Van Kampen, E., "The Structure of a Compact Connected Group," *Amer. Jour. Math.*, **57**, 301–308 (1935).

[3] Pontrjagin, L., "The Theory of Topological Commutative Groups," *Ann. Math.*, **35**, 361–388 (1934).

[4] Malcev, A., "On Solvable Topological Groups (Russian)," *Rec. Math.* (*Mat. Sbornik*), *N.S.*, **19** (61), 165–174 (1946).

[5] Kuranishi, M., "On Locally Euclidean Groups Satisfying Certain Conditions," to appear in *Bull. Amer. Math. Soc.*

[6] Gleason, A., "Spaces with a Compact Lie Group of Transformations," to appear in *Bull. Amer. Math. Soc.*

---

# ON THE AUTOMORPHISMS OF A SFIELD

BY LOO-KENG HUA

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS

Communicated by H. S. Vandiver, May 8, 1949

Let $K$ be a sfield. A mapping $(a \to a^\sigma)$ of $K$ onto itself is called a semi-automorphism if it satisfies

$$(a + b)^\sigma = a^\sigma + b^\sigma \tag{1}$$

$$(aba)^\sigma = a^\sigma b^\sigma a^\sigma \tag{2}$$

and

$$1^\sigma = 1. \tag{3}$$

The well-known examples of semi-automorphisms are automorphisms, which satisfy $(ab)^\sigma = a^\sigma b^\sigma$, and anti-automorphisms, which satisfy $(ab)^\sigma = b^\sigma a^\sigma$.

It is a known problem about the existence of semi-automorphism other than automorphisms and anti-automorphisms. It is the aim of the note to settle this problem, namely:

THEOREM 1. *Every semi-automorphism is either an automorphism or an anti-automorphism.*

To prove the theorem we need several simple consequences of (1), (2) and (3).

Putting $b = a^{-1}$ in (2), we have

$$(a^{-1})^\sigma = (a^\sigma)^{-1}, \tag{4}$$

and replacing $a$ by $a + b$ and $b$ by 1 in (2), we have, by (3),

$$(ab)^\sigma + (ba)^\sigma = a^\sigma b^\sigma + b^\sigma a^\sigma. \tag{5}$$

Applying (2) twice, we obtain, by (4),

$$\begin{aligned}(ba)^\sigma &= (ab(ab)^{-1}ba)^\sigma \\ &= a^\sigma b^\sigma (ab)^{\sigma-1} b^\sigma a^\sigma.\end{aligned} \tag{6}$$

Substituting (6) into (5), we have

$$(ab)^\sigma + a^\sigma b^\sigma (ab)^{\sigma-1} b^\sigma a^\sigma = a^\sigma b^\sigma + b^\sigma a^\sigma$$

which is equivalent to

$$((ab)^\sigma - a^\sigma b^\sigma)(1 - (ab)^{\sigma-1} b^\sigma a^\sigma) = 0. \tag{7}$$

We deduce immediately

$$(ab)^\sigma = \begin{cases} a^\sigma b^\sigma, & \text{or} \\ b^\sigma a^\sigma. \end{cases} \tag{8}$$

Suppose that we have a pair of elements $a$ and $b$ such that

$$(ab)^\sigma = b^\sigma a^\sigma \neq a^\sigma b^\sigma. \tag{9}$$

Then, for any $c$, we have

$$(ac)^\sigma = c^\sigma a^\sigma. \tag{10}$$

In fact, otherwise, we would have, by (8),

$$(ac)^\sigma = a^\sigma c^\sigma \neq c^\sigma a^\sigma,$$

and, by (9) and (1)

$$a^\sigma c^\sigma + b^\sigma a^\sigma = (ac)^\sigma + (ab)^\sigma = (a(b+c))^\sigma = \begin{cases} a^\sigma(b^\sigma + c^\sigma) & \text{or} \\ (b^\sigma + c^\sigma)a^\sigma \end{cases}$$

Both conclusions are impossible. Similarly, we prove that, for any $b$,

$$(db)^\sigma = b^\sigma d^\sigma. \tag{11}$$

Now we are going to prove that

$$(dc)^\sigma = c^\sigma d^\sigma.$$

Suppose the contrary, that is, by (8),

$$(dc)^\sigma = d^\sigma c^\sigma (\neq c^\sigma d^\sigma) \tag{12}$$

Similar to the previous argument, we can establish that

$$(ac)^\sigma = a^\sigma c^\sigma \tag{13}$$

and

$$(db)^\sigma = d^\sigma b^\sigma. \tag{14}$$

Now we consider the elements:

$$b^\sigma a^\sigma + (ac)^\sigma + (db)^\sigma + d^\sigma c^\sigma = ((a+d)(b+c))^\sigma = \begin{cases} (a^\sigma + d^\sigma)(b^\sigma + c^\sigma) & \text{or} \\ (b^\sigma + c^\sigma)(a^\sigma + d^\sigma). \end{cases}$$

The first conclusion contradicts (9), by (10) and (11), and the second conclusion contradicts (12), by (13) and (14).

Therefore, if there is a pair of elements $a,b$ such that $(ab)^\sigma = b^\sigma a^\sigma (\neq a^\sigma b^\sigma)$, then, for any pair of elements $c$ and $d$, we have $(cd)^\sigma = d^\sigma c^\sigma$. This proves our theorem.

By means of the previous theorem, we also settle a problem in the study of projective geometry on a line over a sfield. Namely:

THEOREM 2. *Any one to one mapping carrying the projective line over a sfield of characteristic $\neq 2$ onto itself and keeping harmonic relation invariant is a semi-linear transformation induced by an automorphism or an anti-automorphism.*

This theorem was established for the quaternion algebra (Ancochea[1]) and later for division algebra (Ancochea[2]), characteristic $\neq 2$ and he left the general problem open.

As an easy consequence of the following theorem on geometry of matrices (the proof will be given elsewhere), we can extend theorem 1 to any semi-simple ring with a descending chain condition.

THEOREM 3. *Two $n \times m$ matrices $Z$ and $W$ are said to coherent, if the rank of their difference $Z - W$ is one. Suppose $1 < n \leq m$. Any one to* *one mapping carrying $n \times m$ matrices into $n \times m$ matrices and leaving the coherence relation invariant is of the form*

$$Z_1 = PZ^{\sigma}Q + R \tag{15}$$

*where $P\ (=P^{(n)})$ and $Q(=Q^{(m)})$ are non-singular and $R = R^{(n,m)}$, and $\sigma$ is an automorphism of the sfield. In case $n = m$, in addition to (15), we have also*

$$Z_1 = PZ'^{\sigma}Q + R \tag{16}$$

*where $\sigma$ is an anti-automorphism of the sfield and $Z'$ is the transposed matrix of $Z$.*

By entirely different methods, Ancochea[2] and Kaplansky[3] treated the problem of semi-automorphisms under some restrictions. The former established the corresponding theorem for a simple algebra over a field of characteristic $\neq 2$, and the later extended it to semi-simple algebra over any field. Both of their methods rooted in the structure theory of linear algebras, therefore neither of them can be extended to the general case.

---

[1] Ancochea, G., *J. Math.*, **184**, 192–198 (1942).

[2] Ancochea, G., *Annals Math.*, **48**, 147–153 (1947).

[3] Kaplansky, I., *Duke Math. J.*, **14**, 521–525 (1947).

---

# *LIE RINGS IN PATH SPACE*

BY D. D. KOSAMBI

TATA INSTITUTE OF FUNDAMENTAL RESEARCH IN BIOLOGY, BOMBAY

Communicated by Oswald Veblen, May 8, 1949

*1.* In studying the tensor analysis of the system of differential equations

$$\ddot{x}^i + \alpha^i(x, \dot{x}, t) = 0; \quad i = 1, 2, \ldots, n; \quad \dot{x}^i = dx^i/dt, \text{ etc.,} \tag{1.1}$$

their equations of variation

$$\theta u^i \equiv \ddot{u}^i + \alpha^i{}_{;r}\dot{u}^r + \alpha^i{}_{,r}u^r = 0 \tag{1.2}$$

have been found to be a prime tool of exploration. The notation used is, for any function $\varphi(x, \dot{x}, t)$,

$$\varphi_{,r} = \frac{\partial\varphi}{\partial x^r}; \quad \varphi_{;k} = \frac{\partial\varphi}{\partial\dot{x}^k}; \quad \dot{\varphi} = \frac{d\varphi}{dt} = \frac{\partial\varphi}{\partial t} + \varphi_{,r}\dot{x}^r - \varphi_{;r}\alpha^r \tag{1.3}$$

and the tensor summation convention is followed for indices repeated in subscript and superscript.

These equations of variation (1.2) to be regarded as partial differential equations in $u^i$ are obtained from (1.1) by the "infinitesimal change," $\bar{x}^i = x^i + u^i\delta\tau$, where $\bar{x}$ and $x$ are supposed to be coördinates of points on "nearby" paths. Thus, $\delta x^i = u^i\delta\tau$ and the convention usually adopted is $\delta\dot{x}^i = \dot{u}^i\delta\tau$, which is derived from the assumption $d\delta - \delta d = 0$; this appears in all classical texts without further clarification. The $\alpha^i$ are assumed to be arbitrarily differentiable, which allows an expansion in which second and higher powers of the parameter $\delta\tau$ are to be neglected.

*2.* The only meaning (unless we fix the basic path in 1.2) that can be attached to $\delta x^i = u^i\delta\tau$ is of an infinitesimal transformation of a one-parameter Lie group. But the solutions of $\theta u^i = 0$ are generally functions of $x, \dot{x}, t$, whence the most general such one-parameter group must be taken as operating in the $V_{2n+1}$ of $t$, $x$, $\dot{x}$, where $\dot{x}^i$ is to be regarded as a fiber-bundle attached to a generic point of the base-space defined by coördinates $x^i$. The only operator which can be used here must necessarily be of "the first extension," i.e.,

$$X \equiv u^r \frac{\partial}{\partial x^r} + \dot{u}^r \frac{\partial}{\partial \dot{x}^r} \tag{2.1}$$

and the usual infinite series expansion

$$\bar{\psi} = \psi + \tau X\psi + \frac{\tau^2}{2} X^2\psi + \dots \tag{2.2}$$

can then be obtained, at least symbolically, in the case where the $\alpha^i$ are analytic, to which we restrict our entire discussion. Given several distinct solutions $u^i$, $v^i$, $w^i$, ... of the equations of variation with the associated (extended) operators $X$, $Y$, $Z$, ... the question then naturally arises as to the existence of a Lie group, with more than one parameter being formed in some way out of these. For this, the alternant $\{X, Y\} = XY - YX$ must itself be an operator of the first extension with respect to the paths. That is, if

$$XY - YX = \mu^r \frac{\partial}{\partial x^r} + \lambda^r \frac{\partial}{\partial \dot{x}^r}, \text{ then } \dot{\mu}^r - \lambda^r = 0. \tag{2.3}$$

Furthermore, the Jacobi condition must also be satisfied:

$$\{X, \{Y, Z\}\} + \{Y, \{Z, X\}\} + \{Z, \{X, Y\}\} = 0 \quad \text{for all } X, Y, Z. \tag{2.4}$$

After this, we can see whether the solutions of the equations of variation form a Lie algebra, and over what fields.

Direct calculation gives us

THEOREM 1. *Two (extended) operators $X$, $Y$ associated with vectors $u^i$, $v^i$ alternate to give one of the same type if the vectors concerned are (each) solutions* *of the equations of variation* $\theta u^i = 0$ *or of* $u^i_{;j} = 0$. *Similarly for the Jacobi condition on three operators.*

The proofs can be shortened considerably by noting another result

THEOREM 2. *The solutions of* $\theta u^i = 0$ *are just those whose associated operators permute with the linear operator* $d/dt$ *of 1.3.*

*Proof:* We have

$$Xd/dt - dX/dt \equiv \theta u^r \frac{\partial}{\partial \dot{x}^r}, \qquad (2.5)$$

which suffices for the second theorem. To use this result, we may put $Z = XY - YX$, and note that

$$dZ/dt - Zd/dt = dXY/dt - dYX/dt - XYd/dt + YXd/dt \equiv 0. \quad (2.6)$$

The vanishing identically follows from the permutability of $d/dt$ with both $X$ and $Y$, and then proves that $XY - YX$ is also permutable with $d/dt$, whence the associated operator for $Z$ must be formed from the solutions of the equations of variation, provided of course it was of the first extension. For solutions of $u^i_{;j} = 0$, the proof is trivial. Similarly for the Jacobi identity.

THEOREM 3. *The solutions of* $\theta u^i = 0$ *form a vector space which gives a Lie ring over the set of functions* $\varphi$ *with* $\dot{\varphi} = 0$ *(i.e., constant along any path) and a Lie algebra (and therefore defines a Lie group in the analytic case under discussion) over the field of all real constants.*

The latter statement follows from well-known results in Lie groups. For the former case, we have merely to note that $\varphi u^i$ is a solution of the equations of variation with $u^i$ provided $\dot{\varphi} = 0$. However, we only have here $\{\varphi X, Y\} = \varphi\{X, Y\} - (Y\varphi)X$, whence we get a *ring* over the set $\dot{\varphi} = 0$ defined by the vector space of solutions of $\theta u^i = 0$. The basis of the ring is clearly of dimension $2n + 1$. For the Lie group, however, if we take the most general case, the dimension cannot be finite, and the question remains open whether the infinite Lie algebra and group thus obtained are equivalent to E. Cartan's infinite Lie groups.

*3.* We now consider the subgroups (over the field of all real constants) leaving the base space of the $x$ as well as the paths invariant; the generators must now satisfy $\partial u^i/\partial t = 0$; $u^i_{;j} = 0$; $\theta u^i = 0$. This leads to

THEOREM 4. *The Lie group leaving the base space as well as the paths invariant is of order* $\leq n(n + 1)$.

It suffices to show that the total number of arbitrary parameters in the solutions is finite, $\leq n(n + 1)$ for then these can be specialized to give that number of independent basic solutions and the linearity of the equations allows a general solution to be formed out of linear combinations of these basic solutions. That is, the number of essential parameters being determined, they can be taken to occur in the linear combinations alone.

In this case, we have $\dot{u}^i = u^i_{,r}\dot{x}^r$ and $\ddot{u}^i = u^i_{,j,k}\dot{x}^j\dot{x}^k u^i_{,r}\alpha^r$. Now the equations $\theta u^i = 0$ may be differentiated successively, because of $\partial u^i/\partial t = 0$ to give a succession of homogeneous linear conditions on $u^i_{,j}$, $u^k$. Differentiating the equations of variation with respect to $\dot{x}^s$ successively gives on the *second* differentiation an explicit equation for $u^i_{,j,k}$, and thereafter linear homogeneous restrictions in $u^i$, $u^i_{,r}$. From the original equations $\theta u^i = 0$, and from the first $\dot{x}$ derivative thereof, we may therefore eliminate $u^i_{,j,k}$, and obtain two more linear homogeneous restrictions on $u^i_{,r}, u_j$. The problem therefore is reduced to solving a system of first order partial differential equations for the variables $u^i$ and $v^i_j = u^i_{,j}$ (adjoining this last differential equation to the system), along with linear homogeneous restrictions upon the variables $v^i_j$, $u^k$, to which may be added others derived from the compatibility conditions. In any case, the solution $u^i = v^j_k = 0$ always exists, corresponding to the identity as the sole Lie group for the path-space. But it is well known (cf., E. Goursat, chapter 1 of "Leçons sur . . . équations aux dérivées partielles") that the total number of arbitrary parameters in the general solution is equal to the number of variables which cannot be eliminated from the conditions of compatibility and the linear restrictions, which in no case can exceed the total original number of the variables. Here, that number is $n$ for the $u^i$ plus $n^2$ for the $v^i_j$, proving our theorem.

It is easy to see that *the maximum number of parameters for the group may actually be attained.* The simplest example is of $\ddot{x}^i = 0$, the paths being straight lines. The group is then that of the translations with $n$ parameters, plus the linear transformations leaving the origin invariant, of order $n^2$. In the Riemannian case, for example, as with the equations of Killing, something more is demanded, namely the invariance of a quadratic form as well, whence the maximum order is half the above. For the path-space of straight lines, if we impose, say, a Euclidean metric, the group is then translations plus rotations, order $n + n(n - 1)/2 = n(n + 1)/2$.

*4.* Further extensions of the previous results are possible in several directions, e.g.:

*The group whose generating vectors satisfy only* $u^i_{;j} = 0$, $\theta u^i = 0$, *thus transforming into itself the space* $(x, t)$, *while leaving the absolute parameter* $t$ *unchanged, has a number of parameters* $\leq n(n + 2)$.

The point transformations corresponding to this are Cartan's group B, and their tensor invariants can be found now in an obvious way. The proof parallels the above step by step.

These processes can be carried out also for systems of ordinary differential equations of higher order, as well for partial differential equations, the sole condition being that they be explicitly soluble for the derivatives of highest total order. The $d/dt$ operator for ordinary differential equations has to be again defined as total differentiation along the paths, whereas for partial differentiation we have as many such operators as there are independent variables. We can then prove as before:

THEOREM 5: *The results of theorems 1–3 are valid also for the systems*

$$\frac{d^{\sigma+1}x^i}{dt^{\sigma+1}} + \alpha^i\left(t, x, \frac{dx^i}{dt}, \ldots, \frac{d^{\sigma}x^i}{dt^{\sigma}}\right) = 0, \tag{4.1}$$

*the equations of variation being defined by all those (extended) operators which permute with d/dt, itself defined as:*

$$\frac{d}{dt} \equiv \frac{\partial}{\partial t} + \frac{dx^r}{dt}\frac{\partial}{\partial x^r} + \ldots + \frac{d^{\sigma}x^r}{dt^{\sigma}}\frac{\partial}{\partial x^{(\sigma-1)r}} - \alpha^r\frac{\partial}{\partial x^{(\sigma)r}}. \tag{4.2}$$

*The maximum number of parameters in the group leaving base-space as well as paths invariant cannot then exceed the sum of the first* ($\sigma + 1$) *terms of the series*

$$n\left\{1 + n + \frac{n(n+1)}{2} + \frac{n(n+1)(n+2)}{3} + \ldots\right\}. \tag{4.3}$$

For partial differential equations, we consider only the second order system:

$$\frac{\partial^2 x^i}{\partial t^2 \partial t^{\beta}} + \Pi^i_{\alpha\beta}(t, x, p^j_{\gamma}) = 0; \qquad \begin{aligned} \alpha, \beta &= 1, \ldots m, \\ i, j, \ldots &= 1, \ldots n, \\ p^i_{\alpha} &= \partial x^i/\partial t^{\alpha}. \end{aligned}$$

Here the operators corresponding to $d/dt$ are the set $\partial_{\alpha}$ defined by

$$\partial_{\alpha} \equiv \partial/\partial t^{\alpha} + p^r_{\alpha}\partial/\partial_{,}r - \Pi^r_{\alpha\beta}\partial/\partial p^r_{\beta}. \tag{4.5}$$

THEOREM 6. *The equations of variation have as solutions those vectors and only those whose associated (extended) operators form the ring that permutes with all operators* $\partial_{\alpha}$. *The maximum number of parameters for the group leaving base-space and paths invariant cannot exceed* $n(n + 1)$, *as before.*

The proof is by following the case of ordinary differential equations step by step, and the condition for composition of two operators as well as the Jacobi condition may be derived by direct calculation. The number of parameters is again from considerations of the variation vector $u^i$ being a function of the same number of variables $x$, as is seen directly from the structure of the equations of variation. However, in this case, there are also equations of variation for the independent variables $t^{\rho}$, and it should be made clear that the base-space is of the variables $x^i$.

For equations (1.1) which are deducible from a metric, i.e., the extremals of a regular problem of the calculus of variations, we have the following formula. If the (inverse) Eulerian equations be abbreviated by $\delta_i f = 0$, then the result of any operation $X$ of the base-space group is:

$$\delta_i(Xf) = X(\delta_i f) + u^r_{,i}\delta_r f + f_{;i;r}\theta u^r. \qquad (4.6)$$

This shows what would otherwise have been expected:

THEOREM 7: *If a metric exists for the paths, it is carried into metric by any transformation of the group preserving base space and paths.*

For the simple case $\ddot{x}^i = 0$, a general metric is any arbitrary function $f(x)$ with the determinant $|f_{;i;j}| \neq 0$, and not containing the $x^i$ at all. This is carried into another of the type by any linear homogeneous transformation, and into itself by any translation. The only possible additive terms are necessarily of type $dh/dt$, where $h(x, x)$ is homogeneous of degree zero in $\dot{x}$, as can easily be shown.

It is to be noted, in conclusion, that the conditions of theorem 1 are not necessary. For the alternant of two extended operators associated with vectors $u$, $v$ to give an extended operator the precise condition is

$$v^i_{;r}\theta u^r - u^i_{;r}\theta v^r = 0. \qquad (4.7)$$

Similarly, the Jacobi condition for three such operators is satisfied if and only if

$$u^i_{;r}\{v^r_{;k}\theta w^k - w^r_{;k}\theta v^k\} + (\text{two more terms by cyclic rotation}) = 0. \qquad (4.8)$$

This shows, in particular, that it even suffices to have one of the two vectors a solution of both $\theta u^i = 0$ and $u^i_{;j} = 0$; in the Jacobi condition, it is again sufficient for one of the three to satisfy both these equations. Thus, the infinitesimal transformations not containing $\dot{x}^i$, and in particular the subgroup leaving both base-space and paths invariant, have a special position.

The second remark is about the possibility of defining a Lie differential operator that carries tensors into others of the same type, but of defining it in a manner that can be carried over to more general classes of transformations, such as those over the entire path-space. To this end, we may note that any such transformation may be regarded either as a change of variables, or a charge of coördinates. *The (infinitesimal) Lie operation gives the difference of the (infinitesimal) changes in any geometric object due first to regarding the transformation as a change of variables, and then as a change of coördinates.* Thus, for the tensor $T^i_j$ of weight $p$ under transformations preserving the base-space, we have, when $u^i_{;j} = 0$,

$$LT^i_j \equiv T^i_{j,r}u^r + T^i_{j;r}u^r - T^r_j u^i_{,r} + T^i_r u^r_{,j} + pT^i_j u^r_{,r}. \qquad (4.9)$$

Moreover, there is the infinite series expansion as in (2.2), and we have the tensor carried over into another of the same sort, independently of any connection that may be assigned to the $x$-space. The main definition is obviously extensible for more general transformations.

# *THE FRÉCHET VARIATION AND THE CONVERGENCE OF MULTIPLE FOURIER SERIES*

BY MARSTON MORSE AND WILLIAM TRANSUE*

INSTITUTE FOR ADVANCED STUDY, PRINCETON, NEW JERSEY

Communicated April 30, 1949

1. *The Fréchet Variation.*—The present note is an abstract of results whose publication in full will follow. Let $R^{(\mu)}$ denote a Cartesian space of points $s = [s^{(1)}, \ldots, s^{(\mu)}]$. Let $a$ and $b$ be two points in $R^{(\mu)}$ with $a^{(r)} < b^{(r)}$ $(r = 1, \ldots, \mu)$. Let $J_r$ be an interval for $s^{(r)}$ $(r = 1, \ldots, \mu)$ of one of the forms

$$(a^{(r)}, b^{(r)})(a^{(r)}, b^{(r)}][a^{(r)}, b^{(r)})[a^{(r)}, b^{(r)}]. \tag{1.1}$$

By a *μ-interval* in $R^{(\mu)}$ determined by $a$ and $b$ is meant a Cartesian product

$$I^{(\mu)} = J_1 \times J_2 \times \ldots \times J_\mu. \tag{1.2}$$

By an *r-segment* $Q^{(r)}$ in $R^{(\mu)}$ is meant a Cartesian product of the form (1.2) in which $(\mu - r)$ of the intervals $J_i$ reduce to points while the remaining $r$-intervals $J_i$ are as before. $(0 < r \leqq \mu.)$

*Conditions $\hat{F}$.* Let $g$ be a function mapping a $\mu$-interval $I^{(\mu)}$ in $R^{(\mu)}$ into $R^{(1)}$. Let $Q^{(r)}$ be an arbitrary $r$-segment in $I^{(\mu)}$. If $Q^{(r)}$ is closed the Fréchet variation $P^{(r)}[g, Q^{(r)}]$ of $g$ over $Q^{(r)}$ is defined by an obvious extension of the definition of Fréchet. See Fréchet or MT (1). If $Q^{(r)}$ is not closed $P^{(r)}[g, Q^{(r)}]$ is taken as the sup $P^{(r)}[g, H^{(r)}]$ for all closed $r$-segments $H^{(r)} \subset Q^{(r)}$. We say that $g$ satisfies conditions $\hat{F}$ over $Q^{(r)}$ if $P^{(r)}[g, Q^{(r)}] < \infty$, and if there is an $m$-section $H^{(m)}$ of $Q^{(r)}$ by an $m$-plane parallel to an arbitrary $m$-face of $Q^{(r)} (0 < m < r)$ such that $P^{(m)}[g, H^{(m)}] < \infty$.

THEOREM 1.1. *If $g$ satisfies $\hat{F}$ over $I^{(\mu)}$ the following is true. The variation $P^{(r)}[g, G^{(r)}]$ is bounded over all r-segments $Q^{(r)} \subset I^{(\mu)}$. If $H^{(\mu)}$ is any open subinterval of $I^{(\mu)}$, then $P^{(r)}[g, Q^{(r)}] \to 0$ as the vertices of $Q^{(r)}$ tend in $H^{(\mu)}$ to any one vertex of $H^{(\mu)}$; moreover $g(s)$ tends to a definite "sector" limit as $s \in H^{(\mu)}$ tends to any vertex of $H^{(\mu)}$. The function $g$ is measurable, and its points of discontinuity lie on a countable number of $(\mu - 1)$-planes parallel to the coördinate planes.* See MT (4) and (6).

Let $U^{(\mu)}$ be the open interval determined by the origin and the point with unit coördinates. Any other open $\mu$-interval could be used in place of $U^{(\mu)}$ with appropriate change of notation.

THEOREM 1.2. *A function $k$ which satisfies $\hat{F}$ over $U^{(\mu)}$ admits an extension $g$ over $\overline{U^{(\mu)}}$ with the following properties. The function $g$ is continuous at each vertex of $\overline{U^{(\mu)}}$. At each point $u$ on any open r-cell $\sigma_r$ of the boundary of $U^{(\mu)}$ $(r = 1, \ldots, \mu - 1)$ $g(u)$ is the limit of $k(s)$ as $s \to u$ with $s \in \overline{U^{(\mu)}} \cap$*

$\pi_{\mu-r}$, *where* $\pi_{\mu-r}$ *is the* $(\mu - r)$*-plane orthogonal to* $\sigma_r$ *at* $s = u$. *Moreover*

$$P^{(r)}[k, Q^r \cap U^{(\mu)}] = P^{(r)}[g, Q^{(r)}] \quad (r = 1, \ldots, \mu)$$

*for each* $r$*-segment* $Q^{(r)}$ *in* $\overline{U^{(\mu)}}$ *intersecting* $U^{(\mu)}$.

2. *Sector Functions.*—At each point $a$ in $R^{(\mu)}$ the $(\mu - 1)$-planes on which $s^{(r)} = a^{(r)} (r = 1, \ldots, \mu)$ divide $R^{(\mu)}$ into $2^\mu$ open regions or sectors. Let $S_a$ be any one of these regions. As $a$ varies over $R^{(\mu)}$ we suppose that $S_a$ remains invariant in orientation and refer to $S$ as a *sector type.* There are $2^\mu$ sector types. Let $g$ be defined over a general $\mu$-interval $I^{(\mu)}$. We say that $g$ is $S$*-continuous* [$\bar{S}$-continuous] at a point $a \in I^{(\mu)}$ if $S_a$ intersects $I^{(\mu)}$, and if $g(s) \to g(a)$ as $s \to a$ in $S_a$ [$\bar{S}_a$].

THEOREM 2.1. *If* $g$ *satisfies* $\hat{F}$ *over* $U^{(\mu)}$ *and is* $S$*-continuous then* $g$ *is* $\bar{S}$*-continuous.*

If $\sigma$ is one of the $r$-cells of the boundary of $U^{(\mu)}$ we say that a sector type $S$ induces a sector type $S^\sigma$ in the $r$-plane of $\sigma$, if the sectors $S_a$ of $R^{(\mu)}$ project orthogonally into the sectors $S^\sigma_b$ in the $r$-plane of $\sigma$.

THEOREM 2.2. *If* $g$ *satisfies* $\hat{F}$ *over* $\overline{U^{(\mu)}}$ *then corresponding to any given sector type* $S$ *in* $R^{(\mu)}$ *there exists a function* $g^S$ *with the following properties. At each point* $s$ *of continuity of* $g$ *in* $U^{(\mu)}$, $g(s) = g^S(s)$ *and* $g^S$ *is* $S$*-continuous over* $U^{(\mu)}$. *At each vertex* $a$ *of* $U^{(\mu)}$, $g(a) = g^S(a)$. *On each open boundary* $r$*-cell* $\sigma$ *of* $U^{(\mu)}$, $g|\sigma$ *and* $g^S|\sigma$ *are related as are* $g$ *and* $g^S$ *on* $U^{(\mu)}$ *(relative to* $S^\sigma$ *continuity on* $\sigma$*). Moreover* $P^{(\mu)}[g^S, \overline{U^{(\mu)}}]$ *is independent of the sector type* $S$*, and at most* $P^{(\mu)}[g, \overline{U^{(\mu)}}]$.

The properties of $g^S$ permit an extension and completion of Fréchet's theorem on bilinear functionals over $C \times C$. We consider the $\mu$-fold Cartesian product $C^\mu$ of the Banach space $C$ by itself. Let $f_1, \ldots, f_\mu$ be elements in $C$. An element in $C^\mu$ has the form $[f_1, \ldots, f_\mu] = f$. A functional $B$ over $C^\mu$ with values $B(f_1, \ldots, f_\mu)$ will be termed *multilinear* if $B$ is additive and homogeneous in each argument separately and if there exists a constant $M$ such that for each $f \in C^\mu$

$$|B(f)| \leqq \|f_1\| \ldots \|f_\mu\| M. \tag{2.1}$$

At least constant $M$ such that (2.1) holds for all $f \in C^\mu$ is termed *a minimum modulus* for $B$. A function $g^S$ of type $S$ satisfying $\hat{F}$ over $\overline{U^{(\mu)}}$ and vanishing on the coördinate planes will be termed *canonical of type S.*

THEOREM 2.3. *The most general functional* $B$ *which is multilinear over* $C^\mu$ *has the form of an iterated R-S-integral.*

$$B(f) = \int_0^1 f_1(s^{(1)}) d_{s^{(1)}} \int_0^1 f_2(s^{(2)}) \ldots d_{s^{(\mu-1)}} \int_0^1 f_\mu(s^{(\mu)}) d_{s^{(\mu)}} k(s^{(1)}, \ldots, s^{(\mu)}), \tag{2.2}$$

*where* $k$ *is canonical of type* $S$ *and is uniquely determined by* $B$ *and the sector type* $S$. *Conversely every integral of the form (2.2) with* $g$ *canonical of type*

*S defines a multilinear B. The minimum modulus of B is* $P^{(\mu)}[k, \overline{U^{(\mu)}}]_1$ *The order of integration in (2.2) is immaterial.*

The theorem of Fréchet when $\mu = 2$ is here completed by the evaluation of the minimum modulus, and by the characterization of a unique $k$ as canonical of type $S$.

*3. The Convergence of Fourier Series.*—We have shown that the classical tests (see Gergen) for the convergence of multiple Fourier series generalizing the tests of Jordan, de la Vallée, Poussin, Young, Lebesgue, Pollard, etc., can be made less restrictive by using the Fréchet variation in place of the Vitali variation. A first theorem in $\mu$-space generalizing the Jordan test is as follows.

THEOREM 3.1. *If g satisfies* $\hat{F}$ *over a closed* $\mu$*-interval* $J^{(\mu)}$ *determined by the origin and the point* $[s^{(1)}, \ldots, s^{(\mu)}] = [2\pi, \ldots, 2\pi]$, *and has the period* $2\pi$ *in each coördinate, then the multiple Fourier series for g converges in the sense of Pringsheim to the mean of the* $2^\mu$ *sector limits of g at a. If g is continuous in addition, the convergence is uniform.*

In establishing this theorem the second law of the mean is replaced as follows. Set $g[s^{(1)}, \ldots, s^{(r-1)}, s^{(r)}+, s^{(r+1)}, \ldots, s^{(\mu)}] = g^r(s)$ whenever the limit implied exists. A function $g$ satisfying $\hat{F}$ over $\overline{U^{(\mu)}}$ will be said to have null left boundary values if $g^r(s) = 0$ at each boundary point $s$ at which $s^{(r)} = 0$ $(r = 1, \ldots, \mu)$. The basic lemma follows.

LEMMA 3.1. *If g satisfies* $\hat{F}$ *over* $\overline{U^{(\mu)}}$ *and has null left boundary values, and if* $f_i$ *is in L over* $[0, 1]$, $(i = 1, \ldots, \mu)$ *then*

$$\left|\int_0^1 \cdots \int_0^1 f_1(s^{(1)}) \cdots f_\mu(s^{(\mu)}) g(s^{(1)}, \ldots, s^{(\mu)}) ds^{(1)} \cdots ds^{(\mu)}\right| \leqq P^{(\mu)}[g, \overline{U^{(\mu)}}] M_1 \cdots M_\mu,$$

where $M_i = \sup_{a,b} \left|\int_a^b f_i(t)dt\right|$ for $0 < a < b < 1$.

Hardy, Krause, Tonelli, Young, Küsterman and others have given extensions for double Fourier series of the tests for simple series. These tests are reviewed by Gergen who gives a test $L_R$ for double series which includes (as he shows) all of the others listed. We are able to weaken the conditions in these tests by using the Fréchet variation in place of the Vitali variation. It will suffice at this point to show how Gergen's test is altered. Our new test will be denoted by $L_{MT}$.

The tests concern a function $\varphi$ in $L$ over the 2-interval $[0, \pi] \times [0, \pi]$ and give conditions on $\varphi$ that the Dirichlet integral

$$\int_0^\pi \int_0^\pi \varphi(s, t) \sin\left(\frac{\pi s}{x}\right) \csc\frac{s}{2} \sin\left(\frac{\pi t}{y}\right) \csc\frac{t}{2}\, ds\, dt = o(1)$$

for principal variables $x > 0$, $y > 0$. Our test $L_{MT}$ includes the test $C$. of $L_R$ according to which

$$\Phi(x, y) = \int_0^x \int_0^y \varphi(s, t) ds\, dt = o(xy) \qquad [C_1]. \qquad (3.1)$$

One sets

$$\begin{aligned} \Delta_x\varphi(u, v) &= \varphi(u + x, v) - \varphi(u, v) \\ \Delta_y\varphi(u, v) &= \varphi(u, v + y) - \varphi(u, v) \\ \Delta_{xy}\varphi(u, v) &= \varphi(u + x, v + y) - \varphi(u, v + y) - \varphi(u + x, y) + \varphi(u, v). \end{aligned}$$

To the variables $x$, $y$ is added a variable $k > 0$. Relative to these variables, orders $\bar{o}(1)$, $\bar{o}(x)$, $\bar{o}(y)$ are defined as by Gergen. Let the Fréchet variation of a function $F$ defined over a 2-interval $[a, a'] \times [b, b]$ be denoted by $P\begin{pmatrix} a'b' \\ a\, b \end{pmatrix}(F)$. Gergen's conditions $L_{R'}$ and $L_{R''}$ are replaced by the following set of conditions

$L_{MT'}$:

$$P\begin{pmatrix} \pi - x, \pi - y \\ kx, ky \end{pmatrix}\left[\overline{\frac{\Delta_{xy}\varphi}{uv}}\right] = \bar{o}(1)$$

$L_{MT''}$:

$$P\begin{pmatrix} x, \pi - y \\ o, ky \end{pmatrix}\left[\overline{\frac{\Delta_y\varphi}{v}}\right] = \bar{o}(x)$$

$$P\begin{pmatrix} \pi - x, y \\ kx, o \end{pmatrix}\left[\overline{\frac{\Delta_x\varphi}{u}}\right] = \bar{o}(y).$$

Although Gergen's conditions are not so written it will be seen that his conditions are obtained if $P$ is replaced by $V$ in the above, where $V$ stands for the Vitali variation.

The other types of test, Jordan's, Young's, Pollard's, de la Vallée Poussin's, etc., are altered by the introduction of $P$ instead of $V$ and are examined. The implications between these tests remain essentially unaltered although the tests become less restrictive because of the use of $P$ in place of $V$. The use of $P$ instead of $V$ requires a basic alteration in the non-formal parts of the theory. In particular one no longer uses the Jordan decomposition of a function of bounded Vitali variation into the difference $P - N$ of monotone functions. A calculus for the Fréchet variation of indefinite double integrals is developed. The following are significant laws of this calculus.

For any $F$ in $L$ over $[0, \pi] \times [0, \pi] = I$ set

$$\bar{F}(u, v) = \int_0^u\int_0^v F(s, t)ds\, dt \qquad [(u, v) \in I].$$

Let $[a, a'] \times [b, b']$ be any subinterval of $I$ and let $f_1$ and $f_2$ be continuous over $[a, a']$ and $[b, b']$, respectively. Then

$$P\begin{pmatrix} a', b' \\ a, b \end{pmatrix}\overline{(f_1 f_2 F)} \leqq \max_{(a \leqq u \leqq a')}.|f_1(u)| \max_{(b \leqq v \leqq b')}.|f_2(v)| \; P\begin{pmatrix} a', b' \\ a, b \end{pmatrix}(\bar{F}).$$

If $f$ is of bounded Jordan variation over $[b, b']$ with values $f(v)$, integration by parts as to $v$ of an integral

$$\int_a^{a'}\int_b^{b'}|f(v)F(u,v)|\,du\,dv$$

is replaced by the following:

$$P\begin{pmatrix}a', b'\\ a, b\end{pmatrix}\overline{(fF)} \leqq |f(b')|\,P\begin{pmatrix}a', b'\\ a, b\end{pmatrix}(\bar{F}) + \int_b^{b'} P\begin{pmatrix}a', v\\ a, b\end{pmatrix}(\bar{F})\,|df(v)|.$$

Finally, if $a' - a = x$ and $0 \leqq a_0 \leqq a$

$$\left|\int_a^{a'}\int_b^{b'} F(u,v)du\,dv\right| \leqq P\begin{pmatrix}a', b'\\ a, b\end{pmatrix}(\bar{F})$$

$$\leqq P\begin{pmatrix}a, b'\\ a_0, b\end{pmatrix}\overline{(\Delta_x F)} + P\begin{pmatrix}a_0 + x, b'\\ a_0, b\end{pmatrix}(F).$$

A similar law holds with the rôles of $x$ and $y$ interchanged. The last formula applies to the processes logically and historically descended from the Lebesgue test.

* The contribution of Dr. Transue to this paper is in partial fulfilment of a contract between the Office of Naval Research and the Institute for Advanced Study.

Fréchet, M., "Sur les fonctionelles bilinéaires," *Trans. Am. Math. Soc.*, **16**, 215–244 (1915).

Gergen, J. J., "Convergence Criteria for Double Fourier Series," *Ibid.*, **35**, 29–63 (1933).

Morse, M., and Transue, W.: (1) "Functionals of Bounded Fréchet Variation," *Can. J. Math.*, **1**, 153–165 (1949). (2) "Functionals $F$ Bilinear over the Product $A \times B$ of Two $p$-Normed Vector Spaces. I. The Representation of $F$," *Ann. Math.*, **50** (1949). (3) "Functionals $F$ Bilinear over the Product $A \times B$ of Two $p$-Normed Vector Spaces. II. Admissible Spaces $A$," *Ibid.*, **51** (1950). (4) "Integral Representations of Bilinear Functionals," PROC. NAT. ACAD. SCI., **35**, 136–143 (1949). (5) "A Characterization of the Bilinear Sums Associated with the Classical Second Variation," *Ann. Matematica*, **29** (1950). (6) "The Fréchet Variation in the Small, Sector Limits, and Left Decompositions" (to be published). (7) "The Fréchet Variation and a Generalization of the Jordan Test for Multiple Fourier Series" (to be published). (8) "The Fréchet Variation and Convergence Criteria for Double Fourier Series" (to be published). (9) "Norms of Distribution Functions Associated with Bilinear Functionals" (to be published).

---

# *ON THE ZEROS OF SUCCESSIVE DERIVATIVES OF ENTIRE FUNCTIONS*

BY HANS RÅDSTRÖM

INSTITUTE FOR ADVANCED STUDY, PRINCETON, NEW JERSEY

Communicated by Hermann Weyl, May 31, 1949

In the reference given below I have proved (Theorem 5.4) that given a power series around the origin with finite radius of convergence and such that the origin is not a limit point of the zeros of the successive derivatives it is possible to change the arguments of the coefficients without changing their moduli in such a way that for the new series the origin is such a limit point. On the other hand it is easy to see that the theorem cannot hold for all entire functions. The function $e^z$ is a counter-example. Namely, consider a function $\omega_0 + R(z) = \omega_0 + \omega_1 z/1! + \omega_2 z^2/2! + \ldots$ where $|\omega_\nu| = 1$; $\nu = 0, 1, 2, \ldots$. Obviously $R(z) \leq e^{|z|} - 1$. Thus, if $|z| < \log 2$ it follows: $e^{|z|} - 1 < 1$ or $|R(z)| < |\omega_0|$ and therefore $|\omega_0 + R(z)| > 0$ so that no function of the type considered can have a zero in the circle $|z| < \log 2$. But if one function is of this type then all its derivatives are. The object of this note is to extend the theorem quoted above to all entire functions of infinite order or of finite order greater than one. This is the best possible result with respect to order as the above example shows.

THEOREM: *If* $h(z) = \sum_{\nu=0}^{\infty} b_\nu z^\nu$ *is an entire function of order* $\rho > 1$ *then there exist* $\omega_\nu$ *with* $|\omega_\nu| = 1$ *such that for the function* $k(z) = \sum_{\nu=0}^{\infty} b_\nu \omega_\nu z^\nu$ *the origin is a limit point of the zeros of* $k^{(n)}(z)$.

We assume that the conclusion of the theorem does not hold and we prove that $\rho \leq 1$. We need several lemmas.

LEMMA I: *If* $f(z) = \sum_{\nu=0}^{\infty} a_\nu z^\nu$ *is entire of order* $\rho$ *then*

$$\bar{\theta} = \overline{\lim} \frac{\log \left| \frac{a_{n+1}}{a_n} \right|}{\log n} \geq -\frac{1}{\rho}$$

*Proof:* If $\bar{\theta} \geq 0$ there is nothing to prove because $-\frac{1}{\rho} \leq 0$. Thus, assume $\bar{\theta} < 0$. Let $\theta > \bar{\theta}$ be a negative number. Therefore there is a number $N$ with $\left|\frac{a_{n+1}}{a_n}\right| < n^\theta$ for all $n > N$.

Thus

$$\left|\frac{a_m}{a_{N+1}}\right| = \left|\frac{a_{N+2}}{a_{N+1}}\right| \cdot \left|\frac{a_{N+3}}{a_{N+2}}\right| \cdot \ldots\ldots \cdot \left|\frac{a_m}{a_{m-1}}\right| \leq \left(\frac{m!}{N!}\right)^\theta$$

$$|a_m| < \text{const.} \frac{1}{(m!)^{-\theta}};\ m > N$$

But the order of $\sum_m \frac{z^m}{(m!)^\alpha}$ is $\frac{1}{\alpha}$. Thus $\rho \leq -1/\theta$ or $\theta \geq -1/\rho$ which holds for any $\theta > \bar{\theta}$. Thus $\bar{\theta} \geq -1/\rho$.

LEMMA II: *Let* $P$ *be a polynomial in* $\varphi_1, \varphi_2, \ldots, \varphi_k$, *homogenous of degree* $k$ *and with one term* $= c \cdot \varphi_1^k$. *Then for any given set of values of* $\varphi_1, \varphi_2, \ldots,$

$\varphi_k$ *it is possible to find numbers* $\eta_1\eta_2\ldots\eta_k$ *with* $|\eta_i| = 1$ *such that* $|P(\eta_1\varphi_1, \ldots, \eta_k\varphi_k)| \geq |c\varphi_1^k|$.

*Proof:* Write $P = c\varphi_1^k + Q_1\varphi_1^{k-1} + Q_2\varphi_1^{k-2} + \ldots + Q_k$. Choose $\eta_1 = 1$ and $\eta_i = \eta$, $i = 2, 3, \ldots, k$. Then $A = P(\eta_1\varphi_1, \eta_2\varphi_2 \ldots \eta_k\varphi_k) = c\varphi_1^k + \eta Q_1\varphi_1^{k-1} + \eta^2 Q_2\varphi_1^{k-2} + \ldots \eta^k Q_k$, where $Q_i$ is homogenous of degree $i$ in $\varphi_2, \varphi_3, \ldots, \varphi_k$. Thus $A$ is a polynomial in $\eta$ and because of the maximum modulus theorem $\text{Max}|A(\eta)| \geq |A(0)|$ which is the required inequality for the choice $\eta_1 = 1$; $\eta_2 = \eta_3 = \ldots = \eta_k = \eta_0$ where $\eta_0$ maximizes $|A(\eta)|$ on the circle $|\eta| = 1$.

Now let $f$ be an analytic function such that for $n > N$ no derivative $f^{(n)}(z)$ has a zero in the circle $|z| < r$. Put $f^{(n)}/n! = f_n$; $f_n(0) = a_n$ and $f_{n+1}/f_n = \varphi_n$. Since $f_n$ has no zero within $|z| < r$, $\varphi_n$ is regular there. Of the same reason the multivalued expression $(f_n)^{1/(n+1)}$ represents $n + 1$ different analytic function elements in this circle. For every value of $n$ choose one of these and denote that function by $g_n$. Then $g_n$ is regular and $\neq 0$ in $|z| < r$. We note that $f_n' = (n + 1)f_{n+1}$. Thus: $(n + 1) f_{n+1} = (g_n^{n+1})' = (n + 1)g_n^n \cdot g_n'$, and therefore we get $g_n'/g_n = f_{n+1}/f_n$ or

$$g_n' = g_n\varphi_n \tag{1}$$

We have also

$$\varphi_n' = \left(\frac{f_{n+1}}{f_n}\right)' = \frac{(n + 2)f_{n+2}f_n - (n + 1)f_{n+1}^2}{f_n^2}.$$

or

$$\varphi_n' = (n + 2)\varphi_n\varphi_{n+1} - (n + 1)\varphi_n^2 \tag{2}$$

By means of the formulas (1) and (2) we can, step by step, calculate any derivative of $g_n$. It is easy to verify the following three statements by induction:

1. $g_n^{(k)}$ is a product of $g_n$ and a polynomial $P_{kn}$ in $\varphi_n, \varphi_{n+1}, \ldots, \varphi_{n+k-1}$ with constant coefficients.
2. $P_{kn}$ is homogenous of degree $k$ in its variables.
3. The term containing $\varphi_n^k$ in $P_{kn}$ has the coefficient $c_{kn} = (-1)^{k-1} n(2n + 1)\ldots[(k - 1)n + k - 2]$.

For reference we sum up these facts in a lemma:

LEMMA III: *Let* $f(z)$ *be analytic and have some neighborhood of the origin free from zeros of* $f^{(n)}(z)$; $n > N$. *Then*

$$\frac{g_n^{(k)}(0)}{g_n(0)} = P_{kn}\left(\frac{a_{n+1}}{a_n}; \frac{a_{n+2}}{a_{n+1}}; \ldots; \frac{a_{n+k}}{a_{n+k-1}}\right)$$

*where* $P_{kn}$ *has the above three properties.*

We now proceed to the proof of the theorem. Let $h(z)$ in the theorem be given. If the origin is a limit of zeros of the derivatives, there is nothing to prove. Suppose therefore that the origin is not a limit. Then at most a finite number of the $b_n$ can be equal to zero, i.e., there is an $N$ with $b_n \neq 0$ for $n > N$. We are going to construct the function $k(z)$.

1. Let the following four numbers be given $r_1 > 0$; $\delta > 0$; $p$ a positive integer and $\theta < \bar{\theta} = \overline{\lim} \dfrac{\log \left| \dfrac{b_{n+1}}{b_n} \right|}{\log n}$. Such a number $\theta$ exists for otherwise $\bar{\theta} = -\infty$ which implies $\rho = 0$ (Lemma 1). But $\rho > 1$. Thus there exists a sequence $A$ of natural numbers with

$$\theta < \frac{\log \left| \dfrac{b_{n+1}}{b_n} \right|}{\log n} \text{ for } n \in A \tag{3}$$

2. Consider $l(z) = \sum |b_\nu| z^\nu$. The function $l(z)$ is an entire function with $|k^{(n)}(z)| \leq l^{(n)}(|z|)$ for all $n$, irrespective of the choice of the $\omega_\nu$. Put $\left(\dfrac{l^{(n)}(r_1)}{n!}\right)^{1/(n+1)} = \epsilon_n$. Thus $\epsilon_n \to 0$. Further

$$\left| \frac{k^{(n)}(z)}{n!} \right|^{1/(n+1)} \leq \epsilon_n \text{ for } |z| \leq r_1 \tag{4}$$

3. For every $n > N$ we determine a number (for example, the smallest one) $k_n$ with

$$k_n \geq p \tag{5}$$

and

$$\left( \frac{\epsilon_n}{b_n^{1/(n+1)}} \right)^{\frac{1}{k_n}} < 1 + \delta \tag{6}$$

Obviously, we can choose a sequence $B \subset A$ such that if $n \in B$ then $n + 1$, $n + 2, \ldots, n + k_n$ are not in $B$. Summing up: There exists a sequence $B = \{n_\nu\}$ with $n_{\nu+1} - n_\nu > k_{n_\nu}$ and such that equations (3), (4), (5) and (6) and $n > N$ hold for $n \in B$.

4. Consider $P_{k_n n}\left(\dfrac{b_{n+1}}{b_n}; \ \ldots \ \dfrac{b_{n+k_n}}{b_{n+k_n-1}}\right)$ for a value $n \in B$. According to Lemma III $P$ is homogenous and therefore we can use Lemma II and find multipliers $\eta_i$; $i = n, n + 1, \ldots n + k_n - 1$, such that $|\eta_i| = 1$ and

$$|P_{k_n n}\left(\frac{b_{n+1}}{b_n}\eta_n; \frac{b_{n+2}}{b_{n+1}}\eta_{n+1}; \ \ldots; \frac{b_{n+k_n}}{b_{n+k_n-1}}\eta_{n+k_n-1}\right)| \geq |c_{k_n n}| \cdot \left| \frac{b_{n+1}}{b_n} \right|^{k_n} \tag{7}$$

In this way for every $n \epsilon B$ we get a set of $\eta_i$ determined. It can never happen that for the same $i$ we get two different determinations of $\eta_i$, for we have chosen the density of $B$ sufficiently small. However, we have not necessarily determined $\eta_i$ for every value of $i = 0, 1, 2, \ldots$. For those which still remain undetermined we choose any values subject only to the restriction $|\eta_i| = 1$.

Now put

$$\omega_\nu = \prod_{i=0}^{\nu} \eta_i; \qquad \frac{\omega_{\nu+1}}{\omega_\nu} = \eta_\nu$$

We let these $\omega_\nu$ define $k(z) = \sum_{\nu=0}^{\infty} b_\nu \omega_\nu z^\nu$.

5. Now suppose that there is a number $r_2 > 0$, such that for $n > M$ $k^{(n)}(z)$ has no zero in $|z| \leq r_2$. Then the same is true in $|z| \leq r = \text{Min}\,(r_1; r_2)$. It is therefore possible to apply Lemma III to $k(z)$. We use the same symbols for $k(z)$ as were used in Lemma III for $f(z)$. According to (4) we have $|g_n(z)| \leq \epsilon_n$ for $|z| \leq r$. Cauchy's formulas then give $\left|\frac{g_n^{(k_n)}(0)}{k!}\right| \leq \frac{\epsilon_n}{r^k}$. Further $|g_n(0)| = |b_n|^{1/(n+1)}$.

Thus equation (6) gives

$$\left|\frac{1}{k_n!}\,\frac{g_n^{(k_n)}(0)}{g_n(0)}\right| < \left(\frac{1+\delta}{r}\right)^{k_n} \text{ for } n > M \qquad (8)$$

6. On the other hand, for $n > M$ and $n \epsilon B$ the expression $\left|\frac{g_n^{(k)}(0)}{g_n(0)}\right|$ is equal to the left member of (7):

$$\left|\frac{g_n^{(k_n)}(0)}{g_n(0)}\right| > |c_{k_n n}| \left|\frac{b_{n+1}}{b_n}\right|^{k_n}.$$

We observe that $|c_{kn}| \geq n^{k-1}(k-1)!$ Combining with (8) we get

$$\left|\frac{b_{n+1}}{b_n}\right|^{k_n} n^{k_n-1} < k_n \left(\frac{1+\delta}{r}\right)^{k_n}; \qquad \text{or}$$

$$\frac{\log\left|\frac{b_{n+1}}{b_n}\right|}{\log n} + \frac{k_n - 1}{k_n} < \frac{\log \frac{1+\delta}{r}}{\log n} + \frac{\log k_n}{k_n \log n}$$

But $k_n \geq p$ (5): $\frac{k_n - 1}{k_n} \geq \frac{p-1}{p}$

$$\frac{\log\left|\frac{b_{n+1}}{b_n}\right|}{\log n} < -\frac{p-1}{p} + \frac{\log \frac{1+\delta}{r}}{\log n} + \frac{\log k_n}{k_n \log n}$$

This holds for $n > M$; $n \epsilon B$, i.e., for a subsequence of $A$. But (3) gives

$$\theta < \frac{\log \left| \frac{b_{n+1}}{b_n} \right|}{\log n}.$$

Thus

$$\theta < -\frac{p-1}{p} + \frac{\log \frac{1+\delta}{r} + \frac{\log k_n}{k_n}}{\log n}$$

for an infinity of $n$, or $\theta \leq -(p-1)/p$. But this holds for all $\theta < \bar{\theta}$ and all positive integers $p$.

Let $\theta \rightarrow \bar{\theta}$ and $p \rightarrow \infty$. Thus $\bar{\theta} \leq -1$. On the other hand Lemma I gives $\bar{\theta} \geq -1/\rho$ and we get $\rho \leq 1$. But $\rho > 1$, hence the assumption in 5 is false, which proves the theorem.

Dr. A. Dvoretzky has suggested that the conclusion of the theorem might be true even if the $\omega_\nu$ are permitted to take only the values $+1$ or $-1$. Concerning this problem I have found the following:

1. The "plus-minus" theorem is true for non-entire functions and for entire functions of order $\rho = \infty$ or $\rho > 2$.
2. It is of course not true for $\rho \leq 1$.
3. The above method of proof cannot be used without a deeper study of the polynomials $P_{kn}$, because the "plus-minus" version of Lemma II does not hold if $k \geq 2$. Counter-example for $k = 2$: $P = \varphi_1^2 - \varphi_2^2$ and the given set of values $\varphi_1 = \varphi_2 = 1$.
4. The proof of the "plus-minus" theorem for $\rho > 2$ succeeds by using a method very much the same as the above method. The difference is that I do not let the numbers $k_n$ tend to infinity but give them all the value 2 which makes it possible to use only the polynomial $P_{2n} = \varphi_n[(n+2)\varphi_{n+1} - n\varphi_n]$ for which the "plus-minus" version of Lemma II happens to hold.
5. If the "plus-minus" version of Lemma II can be shown to hold for $P_{kn}$, $k \geq 2$ then the "plus-minus" theorem can be proved for $\rho > k/(k-1)$.

Rådström, H., *Arkiv för Matematik*, 1, No. 7 (1949).

# HOMOTOPY GROUPS OF CERTAIN ALGEBRAIC SYSTEMS

By P. A. Smith

Columbia University

Read before the Academy April 25, 1949

We shall describe an algebraic scheme giving rise to a group $p_2$ that can be identified as the second homotopy group when the elements entering the scheme are suitably interpreted. We consider particularly the group $p_2(X)$ associated with an abstract local group X.

Let $G$ be a groupoid.[1] We denote by $i_l(g)$ and $i_r(g)$ the left- and right-identities of elements $g$ in $G$. We call $g$ closed if $i_l(g) = i_r(g) = i(g)$. We choose a fixed identity $g_0$ in $G$ and denote by $G_0$ the group of closed elements $g$ such that $i(g) = g_0$.

Let $V$ be a set of closed elements of $G$ together with their inverses and identities. For $v \in V$, let $G(v)$ consist of those elements $g$ such that $i_l(g) = g_0$, $i_r(g) = i_l(v)$. Let $A = A(G, V)$ be the totality of formal expressions

$$a = (g_1 * v_1) \dots (g_n * v_n)$$

where $n \geqq 1$, $v_i \in V$, $g_i \in G(v_i)$. We introduce an associative multiplication into $A$ by juxtaposition and a right- and left-neutral element $a_0$. We also introduce an inversion $J$ (automorphism of period 2) by the rule

$$a^J = (g_n * v_n^{-1}) \dots (g_1 * v_1^{-1}).$$

Noting that $hg \in G(v)$ whenever $h \in G_0$, $g \in G(v)$ we see that the elements $g$ of $G_0$ operate on $A$ according to the formula

$$h * a = (hg_1 * v_1) \dots (hg_n * v_n).$$

We agree that $h * a_0 = a_0$.

Let $\beta$ be the mapping $A \rightarrow G_0$ defined as follows:

$$\beta(a) = g_1v_1g_1^{-1} \dots g_nv_ng_n^{-1}.$$

It is evident that $\beta(A)$ is a normal subgroup of $G_0$. Let

$$p_1(A) = G_0/\beta A.$$

We shall see that in certain situations $p_1$ can be identified as the first homotopy group. In order to define $p_2$ we introduce an equivalence relation into $A$. We require this relation to be multiplicative in the sense that $a \sim b$ always implies $a_1aa_2 \sim a_1ba_2$. The generating relations are:

$$g * i(v) \sim a_0 \; (g \in G(v)) \tag{1}$$

$$a_1a_2a_1^J \sim \beta(a_1) * a_2 (a_i \in A) \tag{2}$$

$$(g * v_1)(g * v_2) \sim g * v_1v_2 \tag{3}$$

$$gh * v \sim g * hvh^{-1}. \tag{4}$$

(3) and (4) represent equivalence relations only when the terms are meaningful. Thus (3) is a relation only when $g \epsilon G(v_1) \cap G(v_2)$, $v_1$ and $v_2$ being elements of $V$ such that $v_1v_2 \epsilon G$. (4) is a relation only when $v$, $g$ and $h$ are such that $v \epsilon V$, $hvh^{-1} \epsilon V$, $gh \epsilon G(v)$.

The equivalence classes of $A$ form a multiplicative system $\mathfrak{A}$. It follows from (1) and (3) that $aa^J \sim a_0$ and consequently $\mathfrak{A}$ is a group. It is easy to see that $\beta$ induces a homomorphism

$$\gamma : A \rightarrow \beta(A).$$

Let $p_2(A)$ be the kernel of $\gamma$. The group $p_2(A)$ is contained in the center of $\mathfrak{A}$. For suppose that $a$ is the representative in $\mathfrak{A}$ of an element of $p_2(A)$. Then $\beta(a) = g_0$ so that $aa_1a^J \sim g_0 * a_1 = a_1$ for each element $a_1 \epsilon A$.

*Application to Complexes.*—Let $S$ be a set. We form words with the elements of $S$ as letters and denote their totality, including the empty word, by $W(S)$. If $S$ is a set with inversion $J$, two words of $W(S)$ are called word-equivalent if one can be obtained from the other by deletions and insertions of syllables of the form $ss^J$, $s \epsilon S$. The word-equivalence classes form a group which we denote by $F(S)$.

Now let $K$ be a connected abstract simplicial complex and let $S$ be the totality of the oriented 1-simplexes. Let $s^J$ be the element $s$ of $S$ with orientation reversed. The *paths* of $K$ are representable as elements of $W(S)$ and the classes of word-equivalent paths form a groupoid in a well-known manner. Let $G(K)$ be this groupoid and let $g_0$ correspond to a fixed vertex of $K$. Define $V(K)$ as follows: an element $g$ of $G(K)$ belongs to $V(K)$ if and only if there is a simplex $\sigma_g$ of $K$ such that some path representing $g$ is composed of 1-simplexes of $\sigma_g$. Let

$$A(K) = A(G(K), V(K))$$
$$p_n(K) = p_n(A(K)) \qquad (n = 1, 2).$$

Let $K^*$ be the universal covering complex of $K$ and let $H_n(K)$ denote the $n$-dimensional integral homology group of $K$ based on finite chains.

THEOREM 1. *$p_1(K^*)$ reduces to the identity; $p_2(K) \cong p_2(K^*) \cong H_2(K^*)$. If $K$ corresponds to a finite simplicial decomposition of a polytope, $p_n(K) = \pi_n(K)$, $n = 1, 2$.*

*Application to Local Groups.*—Let $E$ be a set admitting a composition function $e_1 \cdot e_2$ for certain pairs $e_1$, $e_2$ such that when defined, $e_1 \cdot e_2$ is an element of $E$, and such that

1. There exists a unique element 1 such that $1 \cdot e$ and $e \cdot 1$ are defined and equal to $e$ for every $e \epsilon E$;

2. There exists a unique inversion $e \rightarrow e^{-1}$ of $E$ such that $e \cdot e^{-1}$ is defined and equal to 1 for every $e$;

3. If $e_1 \cdot e_2$ is defined, so is $e_2^{-1} \cdot e_1^{-1}$;

4. If $e_1 \circ e_2$, $e_2 \circ e_3$ are defined and if one of the products $(e_1 \circ e_2) \circ e_3$, $e_1 \circ (e_2 \circ e_3)$ is defined, so is the other and the two products are equal.

$E$ is called a local group. Let $w$ be a non-empty element of $W(E)$. By means of the composition function $e_1 \circ e_2$ it may be possible to reduce $w$ to a single element of $E$ by an insertion of parentheses. If every insertion of parentheses is meaningful, we call $w$ a *local word*. Let $L(E)$ be the totality of local words. If $w \in L(E)$, let $\mu(w)$ be the element of $E$ to which $w$ is reducible; it can be shown that $\mu(w)$ is unique.

By a local subgroup of a local group $E$ we mean any symmetric subset of $E$ including the identity, converted into a local group by cutting down the composition function of $E$.

Now let $X$ be a local group and let $G(X)$ be the group $F(X)$, $X$ being regarded as a set with inversion $x \rightarrow x^{-1}$. Let $V(X)$ be defined as follows: an element of $g$ in $G(X)$ belongs to $V(X)$ if and only if $g$ has as representative an element $w$ in $L(X)$ such that $\mu(w) = 1$. Let

$$\begin{aligned} A(X) &= A(G(X), V(X)) \\ p_n(X) &= p_n(A(X)) \qquad (n = 1, 2). \end{aligned}$$

Let $X$ be a local subgroup of a group $Q$. We define a subset $O(X, Q)$ of $F(X)$ as follows: an element $g$ of $F(X)$ belongs to $O(X, Q)$ if and only if $g$ has a representative the product (in $Q$) of whose letters is the identity of $Q$. We call $Q$ *simply connected* relative to $X$ if (1) $X$ generates $Q$ and (2) $O(X, Q) \subset \beta A(X)$.

It is easy to show that the so-called universal covering group of a group $Q$ relative to any local subgroup $X$ which generates $Q$ is simply connected relative to $X$.

Let $X$, $Y$ be local groups with $Y$ a local subgroup of $X$. There can be defined in an obvious manner an inclusion mapping $i$: $p_n(Y) \rightarrow p_n(X)$. Let

$$p_n(Y, X) = ip_n(Y).$$

THEOREM 2. *Let $Q$ be an arcwise connected topological group with the property that for every nucleus $Y$ there exists a nucleus $X$ such that singular 1- and 2-spheres in $Y$ are null-homotopic in $X$. There exist symmetric nuclei $X$ and $Y$ arbitrarily small, such that $p_2(Y, X) = \pi_2(Q)$.*

If $Q$ is a Lie group, it is known[2] that $\pi_2(Q) = 0$. Hence $p_2(Y, X) = 0$ if $X$, $Y$ are suitable symmetric nuclei of a Lie group.

*Extensions of Local Groups.*—Let $E$, $X$ be local groups. Let $\phi$ be a mapping of $E$ onto $X$ and let $\phi_w$ be the mapping $W(E) \rightarrow W(X)$ induced by $\phi$. We call $(E, \phi)$ an extension of $X$ if (1) $\phi_w L(E) = L(X)$, $\phi_w^{-1} L(X) = L(E)$; (2) $\mu\phi(w) = \phi\mu(w)$ whenever $w \in L(E)$.

Let $X$, $Y$ be local subgroups of a group $Q$ with $Y$ a local subgroup of $X$. We call an extension $(E, \phi)$ of $X$ *extensible* over $Q$ from $Y$ if there exists an extension $(T, \psi)$ of $Q$ such that $T$ contains $E$ as local subgroup and $\psi|E = \phi|E$. We call the pair $(Y, X)$ *elementary* with respect to $Q$ if every extension of $X$ is extensible over $Q$ from $Y$.

We can now state a group-theoretic equivalent of the condition $p_2(Y, X) = 0$.

THEOREM 3. *Let $X$, $Y$ be local subgroups of a group $Q$ such that (1) $Y$ is a local subgroup of $X$; (2) $Q$ is simply connected relative to $Y$; (3) $X$ contains no elements of order 2. A necessary and sufficient condition that $(Y, X)$ be elementary with respect to $Q$ is that $p_2(Y, X) = 0$.*

With the aid of Theorem 3, it is possible to give a proof of the existence of Lie groups in the large based on the vanishing of the second homotopy group rather than on the theorem of E. Levi.[3] The details will appear elsewhere.

[1] For definition see Reidemeister, *Einfuhrung in die kombinatorische Topologie*, p. 27.

[2] Although this seems not to be explicitly stated in the literature, it is implied in E. Cartan, *La topologie des groupes de Lie*, p. 13 and pp. 18–23.

[3] See Pontrjagin, *Topological Groups*, p. 269 (theorem 78).

---

# INEQUALITIES BETWEEN THE TWO KINDS OF EIGENVALUES OF A LINEAR TRANSFORMATION

By HERMANN WEYL

THE INSTITUTE FOR ADVANCED STUDY, PRINCETON, NEW JERSEY

Communicated May 15, 1949

With a linear transformation $A$ in an $n$-dimensional vector space (matrix consisting of $n \times n$ complex numbers $a_{ii'}$) there are connected two kinds of eigenvalues: the roots $z = \alpha_1, \ldots, \alpha_n$ of the characteristic polynomial $|zE - A|$ of $A$ ($E$ = unit matrix) and the roots $z = \kappa_1, \ldots, \kappa_n$ of $|zE - K|$ where $K$ is the Hermitian matrix $A^*A$ composed of $A$ and its Hermitian conjugate $A^*$. The $\kappa_i$ are non-negative, and one would naturally compare the $\lambda_i = |\alpha_i|^2$ with the $\kappa_i$. If $A$ is normal, $A^*A = AA^*$, they coincide; in general, however, they do not. Arrange the $\kappa$ as well as the $\lambda$ in descending order,

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n, \qquad \kappa_1 \geq \kappa_2 \geq \ldots \geq \kappa_n.$$

I shall prove the following

THEOREM. *Let $\varphi(\lambda)$ be an increasing function of the positive argument $\lambda$, $\varphi(\lambda) \geq \varphi(\lambda')$ for $\lambda \geq \lambda' > 0$, such that $\varphi(e^{\xi})$ is a convex function of $\xi$ and $\varphi(0) = \lim_{\lambda \to 0} \varphi(\lambda) = 0$. Then the eigenvalues $\lambda_i$ and $\kappa_i$ in descending order satisfy the inequalities*

$$\varphi(\lambda_1) + \ldots + \varphi(\lambda_m) \leq \varphi(\kappa_1) + \ldots + \varphi(\kappa_m) \qquad (m = 1, 2, \ldots, n), \quad (1)$$

*in particular*

$$\lambda_1^s + \ldots + \lambda_m^s \leq \kappa_1^s + \ldots + \kappa_m^s \qquad (m = 1, 2, \ldots, n) \qquad (2)$$

*for any real exponent* $s > 0$.

According to a familiar argument[1]

$$\lambda_1 \leq \kappa_1. \qquad (3)$$

Indeed the equation $Ax = \alpha_1 x$ has a vector solution $x = a \neq 0$: $Aa = \alpha_1 a$, $a^*A^* = \bar{\alpha}_1 a^*$, hence $a^*A^*Aa = \bar{\alpha}_1\alpha_1(a^*a)$ or

$$a^*Ka = \lambda_1(a^*a), \qquad a^*a > 0.$$

Since every vector satisfies the inequality $x^*Kx \leq \kappa_1(x^*x)$, (3) follows.

The linear vector transformation $A$ induces certain linear transformations $A^{[1]}, A^{[2]}, A^{[3]}, \ldots, A^{[n]}$ for the space elements (skew-symmetric tensors) of rank 1, 2, 3, ..., $n$. For instance $A^{[3]} = \|a^{[3]}_{JJ'}\|$ is given by

$$a^{[3]}_{JJ'} = \begin{vmatrix} a_{ii'}, & a_{ik'}, & a_{il'} \\ a_{ki'}, & a_{kk'}, & a_{kl'} \\ a_{li'}, & a_{lk'}, & a_{ll'} \end{vmatrix}$$

where $J$ and $J'$ range over the triples $(i, k, l)$ and $(i', k', l')$ with the restrictions $i < k < l$, $i' < k' < l'$, respectively. Application of the inequality (3) to these matrices $A^{[1]}, A^{[2]}, \ldots$ yields the relations

$$\lambda_1 \leq \kappa_1, \quad \lambda_1\lambda_2 \leq \kappa_1\kappa_2, \quad \ldots, \quad \lambda_1 \ldots \lambda_n \leq \kappa_1 \ldots \kappa_n \qquad (4)$$

(with the equality sign prevailing in the last of them). Everything will be settled as soon as I prove the following

LEMMA: *Let* $\kappa_i, \lambda_i$ $(i = 1, \ldots, m)$ *be non-negative numbers such that*

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m \qquad (5)$$

*and*

$$\lambda_1 \leq \kappa_1, \quad \lambda_1\lambda_2 \leq \kappa_1\kappa_2, \quad \ldots, \quad \lambda_1 \ldots \lambda_m \leq \kappa_1 \ldots \kappa_m; \qquad (6)$$

*then*

$$\textstyle\sum_i \varphi(\lambda_i) \leq \sum_i \varphi(\kappa_i) \qquad (i = 1, \ldots, m) \qquad (7)$$

*for any function* $\varphi$ *of the nature described in the Theorem.*

Of two real numbers $\alpha, \beta$ let max.$(\alpha, \beta)$ denote $\alpha$ if $\alpha \geq \beta$ and $\beta$ if $\beta \geq \alpha$. With a variable argument $z \geq 0$ form the functions

$$f(z) = \prod_{i=1}^{m} \max.(1, \kappa_i z) \quad \text{and} \quad g(z) = \prod_{i=1}^{m} \max.(1, \lambda_i z).$$

Then

$$g(z) \leq f(z) \text{ for } z \geq 0. \tag{8}$$

Indeed set

$$g_i(z) = 1 \text{ for } i = 0 \text{ and } g_i(z) = \lambda_1 \ldots \lambda_i z^i \quad \text{for} \quad i = 1, \ldots, m$$

and distinguish the intervals $\{0\}, \{1\}, \ldots, \{m-1\}, \{m\}$ as defined by

$$\lambda_1 z \leq 1,\ \lambda_1 z \geq 1 \geq \lambda_2 z,\ \ldots,\ \lambda_{m-1} z \geq 1 \geq \lambda_m z,\ \lambda_m z \geq 1.$$

Then $g(z) = g_i(z)$ for $z$ in $\{i\}$. But, because of (6), $g_i(z) \leq f_i(z) \leq f(z)$, hence (8) holds in each of the $m + 1$ intervals.

With an increasing function $\psi(z)$ one can form the Stieltjes integral

$$\int_0^\infty \log g(z) \cdot d\psi(z) = \sum_i \varphi(\lambda_i), \tag{9}$$

provided $\int^\infty \log z \,.\, d\psi(z)$ converges. Here

$$\varphi(\lambda) = \int_0^\infty \log \max.\ (1,\ \lambda z) \cdot d\psi(z) = \int_{\lambda z \geq 1} \log (\lambda z) \cdot d\psi(z). \tag{10}$$

It is clear how (8) by means of (9) and the corresponding formula for $f(z)$ leads to (7).

Set $\lambda = e^\xi$. If $\varphi(\lambda) = G(\xi)$ is a given function satisfying the conditions of the Theorem, it can be expressed by means of a non-decreasing function $G'(t)$ in the form

$$G(\xi) = \int_{-\infty}^{\xi} G'(t) \cdot dt = -\int_{-\infty}^{\xi} (t - \xi) \cdot dG'(t). \tag{11}$$

(The integration per partes is justified since

$$-t \,.\, G'(t) \leq 2 \,.\, \int_t^{t/2} G'(\tau) \cdot d\tau$$

converges to zero for $t \to -\infty$.) (10) goes over into (11) by the substitution $z = e^{-t}$, $\psi(z) = -G'(t)$.

Of the inequalities (2) thus proved, the most important is the last $m = n$, which is independent of any arrangement of the $\kappa_i$ and $\lambda_i$,

$$\lambda_1^s + \ldots + \lambda_n^s \leq \kappa_1^s + \ldots + \kappa_n^s. \tag{2'}$$

Its application to $A^{[2]}, A^{[3]}, \ldots$ yields the further relations

$$\sum_{i_1 < i_2} \lambda_{i_1}^s \lambda_{i_2}^s \leq \sum_{i_1 < i_2} \kappa_{i_1}^s \kappa_{i_2}^s, \tag{2''}$$

$$\sum_{i_1 < i_2 < i_3} \lambda_{i_1}^s \lambda_{i_2}^s \lambda_{i_3}^s \leq \sum_{i_1 < i_2 < i_3} \kappa_{i_1}^s \kappa_{i_2}^s \kappa_{i_3}^s, \tag{2'''}$$

where all the indices $i_1, i_2, i_3, \ldots$ range from 1 to $n$. Together they state that the polynomial $Q_s(z) = \prod_{i=1}^{n} (1 + \lambda_i^s z)$ is majorized, coefficient for coefficient, by the polynomial $P_s(z) = \prod_{i=1}^{n} (1 + \kappa_i^s z)$. In the limit for $s \rightarrow \infty$ they lead back to the relations (4).

If $A$ is non-singular, $A^{-1}$ has the eigenvalues $\alpha_i^{-1}$, and the eigenvalues of $A^{*-1}A^{-1}$ coincide with those of $A^*(A^{*-1}A^{-1})A^{*-1} = A^{-1}A^{*-1} = (A^*A)^{-1}$, i.e., with the $\kappa_i^{-1}$. Hence by application of (1) to $A^{-1}$ corresponding inequalities

$$\sum_{i=m}^{n} \varphi(\lambda_i) \leq \sum_{i=m}^{n} \varphi(\kappa_i) \qquad (m = n, \ldots, 1)$$

will result for any *decreasing* function $\varphi(\lambda)$ for which $\varphi(\lambda) \rightarrow 0$ with $\lambda \rightarrow \infty$ and $\varphi(e^{\xi})$ is convex; in particular for $\varphi(\lambda) = \lambda^s$ with a negative exponent $s$. This shows that for a non-singular $A$ the inequalities (2′) and also (2″), (2‴), ... are valid even for $s \leq 0$.

Facts and proofs, except the last remarks which depend on the consideration of $A^{-1}$, carry over to completely continuous linear operators $A$ in Hilbert space, especially to continuous kernels of integral equations.

Long ago I. Schur proved (2′) for $s = 1$.[2] Recently S. H. Chang showed in his thesis[3] that, in the case of integral equations, convergence of $\sum_i \kappa_i^s$ implies convergence of $\sum_i \lambda_i^s$. These two facts led me to conjecture the relation (2′), at least for $s \leq 1$. After having conceived the simple idea for the proof, I discussed the matter with C. L. Siegel and J. von Neumann; their remarks have contributed to the final form and generality in which the results are presented here.[4]

[1] For a generalization of this inequality see A. Loewy and R. Brauer, "Ueber einen Satz für unitäre Matrizen," *Tôhoku Math. Jour.*, **32**, 44–49 (1930), formula (13) on p. 48.

[2] Schur, I., "Ueber die charakteristischen Wurzeln einer linearen Substitution, mit einer Anwendung auf die Theorie der Integralgleichungen," *Math. Ann.*, **66**, 488–510 (1909).

[3] Chang, S. H., "Theory of Characteristic Values and Singular Values of Linear Integral Equations," Thesis, Cambridge, England, 1948; also, "On the Distribution of Characteristic Values and Singular Values of L² Kernels," *Trans. Am. Math. Soc.* (1949).

[4] While this note was in print a result due to J. Karamata, "Sur une inégalité relative aux fonctions convex," *Publ. Math. Univ. Belgrade*, 1, 145–148 (1932), that comes very near to our lemma, was pointed out to me.

# SEPARATION OF THE VARIABLES IN THE ONE-PARTICLE SCHROEDINGER EQUATION IN 3-SPACE

BY LUTHER PFAHLER EISENHART

FINE HALL, PRINCETON, NEW JERSEY

Communicated May 7, 1949

The one-particle Schroedinger equation in Cartesian coördinates $y_i (i = 1, 2, 3)$ of euclidean 3-space is

$$\sum_i \frac{\partial^2 \psi}{\partial y_i^2} + k^2(E - V)\psi = 0 \qquad (i = 1, 2, 3), \tag{1}$$

where $E$ is the energy and $V$ the potential. For any orthogonal system $x_i$ in the space the fundamental quadratic form of the space is

$$ds^2 = \sum_i H_i^2 \, dx_i^2, \tag{2}$$

where

$$H_i^2 = \sum_j \left(\frac{\partial y_j}{\partial x_i}\right)^2, \qquad \sum_j \frac{\partial y_j}{\partial x_i}\frac{\partial y_j}{\partial x_k} = 0 \quad (i \neq k).$$

For the coördinate system $x_i$ the above equation is

$$\frac{1}{H} \sum_i \frac{\partial}{\partial x_i}\left(\frac{H}{H_i^2}\frac{\partial \psi}{\partial x_i}\right) + k^2(E - V)\psi = 0, \tag{3}$$

where $H$ is the product of the $H_i$'s, that is $H = \prod_i H_i$. If one seeks a solution of this equation of the form

$$\psi = X_1 X_2 X_3, \tag{4}$$

where $X_i$ is a function of $x_i$ alone, the equation may be written

$$\sum_i \frac{1}{H_i^2}\left(\frac{X_i''}{X_i} + \frac{\partial}{\partial x_i} \log \frac{H}{H_i^2}\frac{X_i'}{X_i}\right) + k^2(E - V) = 0, \tag{5}$$

where the primes indicate derivatives.

A necessary and sufficient condition that the quantity in parentheses involve $x_i$ at most is that

$$\frac{H}{H_i^2} = \varphi_i \psi_i, \tag{6}$$

where $\varphi_i$ is a function of $x_i$ alone and $\psi_i$ does not involve $x_i$. Multiplying together the three equations (6) we obtain

$$H = \prod_i \varphi_i \psi_i, \tag{7}$$

and consequently

$$H_1^2 = \varphi_2\varphi_3\psi_2\psi_3, \; H_2^2 = \varphi_3\varphi_1\psi_3\psi_1, \; H_3^2 = \varphi_1\varphi_2\psi_1\psi_2. \tag{8}$$

If the potential energy in the coördinate system $x_i$ is such that

$$k^2(E - V) = \sum_i \frac{f_i}{H_i^2} \tag{9}$$

where $f_i$ is a function of $x_i$ alone, the equation (5) multiplied by $H$ may be written in the form

$$\sum_i \varphi_i\psi_i \left(\frac{X_i''}{X_i} + \frac{\varphi_i'}{\varphi_i}\,\frac{X_i'}{X_i} + f_i\right) = 0, \tag{10}$$

in consequence of (6). Consequently for solutions $X_i$ of the ordinary differential equations

$$X_i'' + \frac{\varphi_i'}{\varphi_i} X_i' + f_i X_i = 0 \tag{11}$$

equation (4) is a solution of equation (3).

Although equation (3) was assumed to be the equivalent of equation (1) expressed in terms of an orthogonal system $x_i$ in euclidean 3-space, the procedure following equation (3) applies to any 3-space whose fundamental quadric form is given by (2) when the $H_i^2$ are of the form (8).

In what follows we derive the forms (8) when the 3-space is euclidean. A first requirement is the Riemann tensor equation $R_{jiik} = 0(i,\ j,k \neq)$, that is[1]

$$2\frac{\partial^2 \log H_i^2}{\partial x_j \partial x_k} + \frac{\partial \log H_i^2}{\partial x_j}\,\frac{\partial \log H_i^2}{\partial x_k} - \frac{\partial \log H_i^2}{\partial x_j}\,\frac{\partial \log H_j^2}{\partial x_k} - \frac{\partial \log H_i^2}{\partial x_k}\,\frac{\partial \log H_k^2}{\partial x_j} = 0$$

From (8) we have

$$\frac{\partial^2 \log H_i^2}{\partial x_j \partial x_k} = 0, \tag{12}$$

so that the condition is

$$\frac{\partial \log H_i^2}{\partial x_j}\,\frac{\partial \log H_i^2}{\partial x_k} - \frac{\partial \log H_i^2}{\partial x_j}\,\frac{\partial \log H_j^2}{\partial x_k} - \frac{\partial \log H_i^2}{\partial x_k}\,\frac{\partial \log H_k^2}{\partial x_j} = 0. \tag{13}$$

If we put

$$\psi_1 = \frac{\alpha_{23}}{\varphi_2\varphi_3}, \quad \psi_2 = \frac{\alpha_{31}}{\varphi_3\varphi_1}, \quad \psi_3 = \frac{\alpha_{12}}{\varphi_1\varphi_2}, \tag{14}$$

where $\alpha_{ij}$ is a function of $x_i$ and $x_j$ at most, we have from (8)

$$H_1^2 = \frac{\alpha_{12}\alpha_{31}}{\varphi_1^2}, \quad H_2^2 = \frac{\alpha_{23}\alpha_{12}}{\varphi_2^2}, \quad H_3^2 = \frac{\alpha_{31}\alpha_{23}}{\varphi_3^2}. \tag{15}$$

For these expressions equations (13) are

$$\begin{aligned}
&\frac{\partial \log \alpha_{12}}{\partial x_2}\left(\frac{\partial \log \alpha_{31}}{\partial x_3} - \frac{\partial \log \alpha_{23}}{\partial x_3}\right) - \frac{\partial \log \alpha_{31}}{\partial x_3}\frac{\partial \log \alpha_{23}}{\partial x_2} = 0,\\
&\frac{\partial \log \alpha_{23}}{\partial x_3}\left(\frac{\partial \log \alpha_{12}}{\partial x_1} - \frac{\partial \log \alpha_{31}}{\partial x_1}\right) - \frac{\partial \log \alpha_{12}}{\partial x_1}\frac{\partial \log \alpha_{31}}{\partial x_3} = 0, \quad (16)\\
&\frac{\partial \log \alpha_{31}}{\partial x_1}\left(\frac{\partial \log \alpha_{23}}{\partial x_2} - \frac{\partial \log \alpha_{12}}{\partial x_2}\right) - \frac{\partial \log \alpha_{23}}{\partial x_2}\frac{\partial \log \alpha_{12}}{\partial x_1} = 0.
\end{aligned}$$

If the first and second of these equations are multiplied by $\frac{\partial \log \alpha_{12}}{\partial x_1}$ and $\frac{\partial \log \alpha_{12}}{\partial x_2}$ respectively and added, we obtain

$$\frac{\partial \log \alpha_{12}}{\partial x_1}\frac{\partial \log \alpha_{23}}{\partial x_2}\frac{\partial \log \alpha_{31}}{\partial x_3} + \frac{\partial \log \alpha_{12}}{\partial x_2}\frac{\partial \log \alpha_{23}}{\partial x_3}\frac{\partial \log \alpha_{31}}{\partial x_1} = 0. \tag{17}$$

The same result follows if the first and third of equations (16) are multiplied by $\frac{\partial \log \alpha_{31}}{\partial x_1}$ and $\frac{\partial \log \alpha_{31}}{\partial x_3}$ respectively and added; also if the second and third are multiplied by $\frac{\partial \log \alpha_{23}}{\partial x_2}$ and $\frac{\partial \log \alpha_{23}}{\partial x_3}$ respectively and added.

We consider first the case when none of the terms in (17) is zero. When the equation is written in the form

$$\frac{\dfrac{\partial \log \alpha_{12}}{\partial x_1}}{\dfrac{\partial \log \alpha_{12}}{\partial x_2}}\,\frac{\dfrac{\partial \log \alpha_{31}}{\partial x_3}}{\dfrac{\partial \log \alpha_{31}}{\partial x_1}} + \frac{\dfrac{\partial \log \alpha_{23}}{\partial x_3}}{\dfrac{\partial \log \alpha_{23}}{\partial x_2}} = 0,$$

it is seen that the first and second parts of the first term do not involve $x_3$ and $x_2$, respectievly, but combined they involve only $x_2$ and $x_3$. If, then, we put

$$\begin{aligned} \frac{\partial \log \alpha_{12}}{\partial x_1} &= a_{12}X_1, & \frac{\partial \log \alpha_{12}}{\partial x_2} &= -a_{12}X_2, \\ \frac{\partial \log \alpha_{23}}{\partial x_2} &= a_{23}X_2, & \frac{\partial \log \alpha_{23}}{\partial x_3} &= -a_{23}X_3, \\ \frac{\partial \log \alpha_{31}}{\partial x_3} &= a_{31}X_3, & \frac{\partial \log \alpha_{31}}{\partial x_1} &= -a_{31}X_1, \end{aligned} \qquad (18)$$

where $X_i$ is a function of $x_i$ alone, and $a_{ij}$ of $x_i$ and $x_j$, equation (17) is satisfied. Substituting these expressions in equations (16), we obtain a single condition on the quantities $a_{ij}$, which may be written in the form

$$\frac{1}{a_{12}} + \frac{1}{a_{23}} + \frac{1}{a_{31}} = 0.$$

Differentiating this equation with respect to $x_1$ and $x_2$, we find that $\frac{1}{a_{12}}$ is the sum of a function of $x_1$ and one of $x_2$; similarly for the other two terms in this equation. Accordingly we have $\frac{1}{a_{ij}} = \sigma_i - \sigma_j$ as the solutions of this equation. Substituting these expressions in the first two of equations (18), we find as the condition of integrability $\frac{X_1}{\sigma_1'} = \frac{X_2}{\sigma_2'}$, that is $X_1 = n\sigma_1'$, $X_2 = n\sigma_2'$, where $n$ is a constant. For the other two of equations (18) we obtain likewise $X_3 = n\sigma_3'$. Accordingly, the solution of (18) is $\alpha_{ij} = (\sigma_i - \sigma_j)^n$, to within a negligible constant factor. From these results and (15) we have that by a suitable choice of $x_i$

$$H_i^2 = X_i(x_i - x_j)^n(x_i - x_k)^n \qquad (i, j, k \neq) \qquad (19)$$

where $X_i$ is a function of $x_i$ alone.

If in the first two of equations (18) $X_1$ and $X_2$ are interchanged, in the second two $X_2$ and $X_3$, and in the third $X_3$ and $X_1$, one again obtains the result (19).

We consider next the case where $\alpha_{12}$ is a function of $x_1$ alone. From (17) it follows that $\alpha_{31}$ is a function of $x_1$ alone or $\alpha_{23}$ of $x_3$ alone. We investigate these two cases.

1°. $\alpha_{12} = \sigma_1$, $\alpha_{31} = \tau_1$, where $\sigma_1$ and $\tau_1$ are functions of $x_1$. The first of equations (16) is satisfied and the other two reduce to

$$\frac{\partial \log \alpha_{23}}{\partial x_3}\left(\frac{\sigma_1'}{\sigma_1} - \frac{\tau_1'}{\tau_1}\right) = 0, \qquad \frac{\partial \log a_{23}}{\partial x_2}\left(\frac{\sigma_1'}{\sigma_1} - \frac{\tau_1'}{\tau_1}\right) = 0.$$

For $\alpha_{23}$ = const. we have from (15) by suitable choice of $x_i$

$$H_1^2 = 1, \qquad H_2^2 = \varphi(x_1), \qquad H_3^2 = \psi(x_1). \qquad (20)$$

For the other solution $\tau_1 = c\sigma_1$ we have

$$H_1^2 = 1, \qquad H_2^2 = H_3^2 = \sigma_1 \alpha_{23}, \tag{21}$$

where $\alpha_{23}$ is an arbitrary function of $x_2$ and $x_3$.

2°. $\alpha_{12} = \sigma_1, \alpha_{23} = \sigma_3$. The first and third of equations (16) are satisfied by these values and the second reduces to

$$\frac{\sigma_3'}{\sigma_3}\left(\frac{\sigma_1'}{\sigma_1} - \frac{\partial \log \alpha_{31}}{\partial x_1}\right) - \frac{\sigma_1'}{\sigma_1}\frac{\partial \log \alpha_{31}}{\partial x_3} = 0,$$

of which the solution is

$$\alpha_{31} = (a\sigma_1 + b\sigma_3) f\left(\frac{\sigma_1}{\sigma_3}\right),$$

where $a$ and $b$ are constants and $f$ is an arbitrary function of $\frac{\sigma_1}{\sigma_3}$. For this case we have

$$H_1^2 = H_2^2 = (a\sigma_1 + b\sigma_3) f\left(\frac{\sigma_1}{\sigma_3}\right), \qquad H_2^2 = \sigma_1\sigma_3. \tag{22}$$

If we take $\alpha_{12} = \sigma_2$, that is a function of $x_2$, and proceed as above, we obtain results equivalent to the above ones. Hence the cases (19), (20), (21) and (22) constitute the possibilities.

We have yet to apply to these cases the requirement $R_{jiij} = 0$, that is[1]

$$\begin{gathered} H_1^2\left(2\frac{\partial^2 \log H_i^2}{\partial x_j^2} + \frac{\partial \log H_i^2}{\partial x_j}\frac{\partial}{\partial x_j}\log\frac{H_i^2}{H_j^2}\right) + H_j^2\left(2\frac{\partial^2 \log H_j^2}{\partial x_i^2} + \right. \\ \left. \frac{\partial \log H_j^2}{\partial x_i}\frac{\partial}{\partial x_i}\log\frac{H_j^2}{H_i^2}\right) + \frac{1}{H_k^2}\frac{\partial H_i^2}{\partial x_k}\frac{\partial H_j^2}{\partial x_k} = 0 \qquad (i, j, k \neq). \end{gathered} \tag{23}$$

For the form (20) by a suitable choice of $x_i$ we have

$$H_1^2 = H_3^2 = 1, \quad H_2^2 = x_1^2. \tag{20a}$$

For (21) we have

$$H_1^2 = 1, \qquad H_2^2 = H_3^2 = (ax_1 + b)^2\alpha_{23}, \tag{21a}$$

where $a$ and $b$ are arbitrary constants and $\alpha_{23}$ is subject to the condition

$$\left(\frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2}\right)\log \alpha_{23} + 2a^2\alpha_{23} = 0. \tag{24}$$

For (22) the conditions (23) for $i = 2, j = 1, 3$ are

$$\sigma_1'^2 = c^2\sigma_1^2 + 2d\sigma_1, \qquad \sigma_3'^2 = -c^2\sigma_3^2 + 2e\sigma_3 \tag{25}$$

where $c$, $d$ and $e$ are constants and $f\left(\frac{\sigma_1}{\sigma_3}\right) = \frac{e\sigma_1 + d\sigma_3}{a\sigma_1 + b\sigma_3}$. Hence we have

$$H_1^2 = H_3^2 = e\sigma_1 + d\sigma_3, \qquad H_2^2 = \sigma_1\sigma_3. \tag{22a}$$

These values satisfy identically equations (23) for $i = 1, j = 3$.

For the expressions (19), the equation (23) for $i = 1, j = 2$ is

$$\frac{1}{(x_2 - x_1)^{n+1}}\left\{(x_1 - x_3)^{n+1}\left[(x_2 - x_3)\left(\frac{1}{X_2}\right)' - \left(\frac{2(x_2 - x_3)}{x_2 - x_1} + n\right)\frac{1}{X_2}\right] + \right.$$
$$\left.(-1)^{n+1}(x_2 - x_3)^{n+1}\left[(x_1 - x_3)\left(\frac{1}{X_1}\right)' - \left(\frac{2(x_1 - x_3)}{x_1 - x_2} + n\right)\frac{1}{X_1}\right]\right\} + \frac{n}{X_3} = 0.$$

When this equation is differentiated with respect to $x_2$, $x_3$ appears in the result either in the form $(x_1 - x_3)^{n+1}x_3$ or $(x_2 - x_3)^{n+1}x_3$. Hence the coefficient of $x_3^{n+2}$ must be equal to zero. This gives the equation

$$(x_1 - x_2)^2\left(\frac{1}{X_2}\right)'' + (n + 3)(x_1 - x_2)\left(\frac{1}{X_2}\right)' + 2(n + 2)\frac{1}{X_2} +$$
$$(-1)^{n+1}\left[(n + 1)(x_1 - x_2)\left(\frac{1}{X_1}\right)' - 2(n + 2)\frac{1}{X_1}\right] = 0. \tag{27}$$

Differentiating this equation once with respect to $x_2$ and twice with respect to $x_1$, we obtain

$$2\left(\frac{1}{X_2}\right)''' = (-1)^{n+1}(n + 1)\left(\frac{1}{X_1}\right)'''.$$

Hence each member of this equation is a constant, say $12a_0$. Accordingly

$$\frac{1}{X_2} = a_0x_2^3 + a_1x_2^2 + a_2x_2 + d_3,\ (-1)^{n+1}(n + 1)\frac{1}{X_1} =$$
$$2(a_0x_1^3 + b_1x_1^2 + b_2x_1 + b_3).$$

where the $a$'s and $b$'s are constants. Substituting in (27) we find that $n = 1, b_i = a_i$. Hence, if we put for $i = 1, 2$

$$f(x_i) = a_0x_i^3 + a_1x_i^2 + a_2x_i + a_3, \tag{28}$$

we have that $1/X_1 = f(x_1)$ and $1/X_2 = f(x_2)$, and from (26) $1/X_3 = f(x_3)$. Consequently (19) is of the form

$$H_i^2 = \frac{(x_i - x_j)(x_i - x_k)}{f(x_i)} \qquad (i, j, x \neq) \tag{19a}$$

In 1927 Robertson[2] considered the equation (5) for an $n$-space and sought the conditions that the $H_i^2$ must satisfy in order that there be a solution of the form $\Pi X_i$ involving $n - 1$ arbitrary constants other than $E$ such that the quantities $X_i$ are solutions of ordinary differential equations. Assuming equations (8) and that a certain matrix be of rank $n$ he proved that it is necessary and sufficient that the quantities $H_i^2$ be of the form derived by Stäckel in 1891 in seeking a solution of the form $\Sigma X_i$ of the Hamilton-Jacobi equation, the quantities $H_i^2$ being subject to a certain condition and that $V$ be of the form $\sum\limits_i \frac{f_i}{H_i^2}$ where $f_i$ is a function of $x_i$ at most.

In 1934 I proved that necessary and sufficient conditions that quantities $H_i^2$ be of the Stäckel form are[3]

$$\frac{\partial^2 \log H_i^2}{\partial x_i \partial x_j} + \frac{\partial \log H_i^2}{\partial x_j}\frac{\partial \log H_j^2}{\partial x_i} = 0,$$

$$\frac{\partial^2 \log H_i^2}{\partial x_j \partial x_k} - \frac{\partial \log H_i^2}{\partial x_j}\frac{\partial \log H_i^2}{\partial x_k} + \frac{\partial \log H_i^2}{\partial x_j}\frac{\partial \log H_j^2}{\partial x_k} + \frac{\partial \log H_i^2}{\partial x_k}\frac{\partial \log H_k^2}{\partial x_j} = 0. \quad (29)$$

Since the forms (20a), (21a), (22a) subject to (25), and (19a) satisfy (12) and (13), the second set of (29) are satisfied. All of the forms satisfy the first set of equations (29) except (21a); in the latter case (21a) is replaced by

$$H_1^2 = 1, \qquad H_2^2 = H_3^2 = (ax_1 + b)(\sigma_2 + \sigma_3) \qquad (21\text{b})$$

where $\sigma_2$ and $\sigma_3$ are functions of $x_2$ and $x_3$ respectively.

In a recent paper[4] there are given the various Stäckel forms of $H_i^2$, corresponding to suitable choices of $x_i$ and of the constants in the above forms, and the relations of $x_i$ and cartesian coördinates of the space. In particular, (20a) appears as I; (21b) subject to (24) as II–V; (22a) subject to (25) as VI–VIII; (19a) as IX, X.

When $\alpha_{23}$ in (21a) is not of the form $\sigma_2 + \sigma_3$ but any other solution of (24) the quantities $H_i^2$ are not of the Stäckel form.

[1] Eisenhart, L. P., *Riemannian Geometry*, Princeton University Press, p. 119.

[2] Robertson, H. P., "Bemerkung über separierbare Systeme in der Wellenmechanik," *Math. Ann.*, **98**, 749–752 (1927).

[3] Eisenhart, L. P., "Separable Systems of Stäckel," *Ann. Math.*, **35**, 284–305 (1934).

[4] Eisenhart, L. P., "Enumeration of Potentials for Which One-Particle Schroedinger Equations Are Separable," *Phys. Rev.*, **74**, 87–89 (1948).

# *PHYSICAL FAMILIES IN CONSERVATIVE FIELDS OF FORCE**

BY EDWARD KASNER AND JOHN DE CICCO

DEPARTMENTS OF MATHEMATICS, COLUMBIA UNIVERSITY, NEW YORK, AND ILLINOIS INSTITUTE OF TECHNOLOGY, CHICAGO

Communicated May 7, 1949

*1. A system* $S_k$ of $\infty^3$ curves in a field of force, where the force vector acting at any point of a given region of the plane depends only on the position of the point, consists of curves along which a constrained motion is possible such that the pressure $P$ is proportional to the normal component $N$ of the force vector. Thus $P = kN$ where $k(\neq -1)$ is the constant factor of proportionality.[1]

There is no loss in generality in assuming that the particle is of unit mass. The speed of the particle is denoted by $v$. The arc length and the radius of curvature of a curve $C$ are represented by $s$ and $r$. The subscript $s$ means total differentiation with respect to $s$, whereas subscripts $x$ and $y$ signify partial differentiations. Finally primes denote total differentiation with respect to $x$.

*The equations of motion in intrinsic form of a system* $S_k$ *of* $\infty^3$ *curves are*

$$v^2 = (k + 1)rN, \qquad vv_s = T, \tag{1}$$

*where T and N are the tangential and normal components of the force vector.*

The important special cases of physical interest of a system $S_k$ are (*a*) dynamical trajectories, $k = 0$; (*b*) general catenaries, $k = 1$; (*c*) generalized brachistochrones, $k = -2$, (*d*) velocity curves, $k = \infty$.

In (1), it is assumed that $k \neq -1$ and $k \neq \infty$. Thus, a velocity system $S_\infty$ which is a limiting case of $S_k$ as $k$ becomes infinite must be given separate treatment.[2] A curve $C$ is a *velocity curve corresponding to the speed* $v_0$ if any particle, projected from any point of $C$ in the direction of $C$ with the speed $v_0$, describes a trajectory which is initially osculated by $C$. Thus *a velocity system* $S_\infty$ *is defined by the equation*

$$v_0^2 = rN. \tag{2}$$

*2.* In a conservative field of force, the work done in traversing a path $C$, namely $W = \int_C T ds$, is independent of the path. Therefore in a conservative field of force, the speed $v$ of a particle at a point $(x, y)$ is given by the energy equation

$$\frac{v^2}{2} = W(x, y) + h, \tag{3}$$

where $h$ is the constant of energy.

The components $T$ and $N$ are given by the equations

$$T = W_x \frac{dx}{ds} + W_y \frac{dy}{ds}, \qquad N = W_y \frac{dx}{ds} - W_x \frac{dy}{ds}. \tag{4}$$

In a conservative field of force, the equations (1) admit a first integral

$$v^2 = (k+1)rN = 2(W+h). \tag{5}$$

Therefore *a system $S_k$, where $k \neq -1$ and $k \neq \infty$, in a conservative field of force, consists of the integral curves of the second order differential equation*

$$y'' = \frac{k+1}{2} \frac{W_y - y'W_x}{W+h} (1 + y'^2). \tag{6}$$

By (2), it is found that *the system $S_\infty$ of velocity curves in a conservative field of force consists of the integral curves of the second order differential equation*

$$y'' = \frac{1}{v_0^2} (W_y - y'W_x)(1 + y'^2). \tag{7}$$

*3.* Let a surface $\Sigma$ of space be mapped conformally upon the given region of the plane. The linear element $d\sigma$ of $\Sigma$ is $d\sigma = \mu(x, y)ds = \mu(x, y)(dx^2 + dy^2)^{1/2}$, where $\mu(x, y) > 0$.

The geodesics of $\Sigma$ are the extremals of the variation problem

$$\int \mu ds = \text{min.} \tag{8}$$

By Euler's equation, the $\infty^2$ geodesics of $\Sigma$ are the integral curves of the second order differential equation

$$y'' = \frac{1}{\mu} (\mu_y - y'\mu_x)(1 + y'^2). \tag{9}$$

The extremals of a variation problem of the form (8) are said to form a *natural family.* Thus any natural family of the plane consists of the integral curves of a second order differential equation of the form (9).

*4. A system $S_k$, where $k \neq -1$ and $k \neq \infty$, in a conservative field of force can be separated into $\infty^1$ natural families, each defined by the energy constant $h$. These are the extremals of the variation problem*

$$\int v^{k+1} ds = \text{min.,} \quad \text{or} \quad \int (W+h)^{(k+1)/2} ds = \text{min.} \tag{10}$$

This follows from (6) and (9).

*A velocity system $S_\infty$ in a conservative field of force can be separated into $\infty^1$ natural families, each defined by the speed $v_0$. These are the extremals of the variation problem*

$$\int e^{W/v_0^2} ds = \text{min.} \tag{11}$$

This [illegible]vident from (7) and (9)

5. Assume that $ds$ is the linear element on a surface $\Sigma$ so that

$$ds^2 = E(x, y)dx^2 + 2F(x, y)dxdy + G(x, y)dy^2, \qquad (12)$$

where $E$, $F$, $G$ are defined over a region of the $(x, y)$ plane such that $EG - F^2 > 0$, $E > 0$, $G > 0$.

If this $ds$ is used, the extremals of (10) define a *system* $S_k$ *on the surface* $\Sigma$ and the extremals of (11) define a *velocity system* $S_\infty$ *on* $\Sigma$. The variation problem (8) defines a *natural family on* $\Sigma$.

*If a surface* $\Sigma_1$ *is represented on the surface* $\Sigma$ *by a point transformation, then all natural families on* $\Sigma_1$ *are depicted as natural families on* $\Sigma$ *if and only if the representation is conformal.*

The linear element $d\sigma$ of $\Sigma_1$ is $d\sigma = \mu(x, y)ds$, where $\mu(x, y) > 0$.

6. A system $S_k$ on the surface $\Sigma$, defined by (10), is represented on the surface $\Sigma_1$ by the variation problem

$$\int \left[\frac{W + h}{\mu^{2/(k+1)}}\right]^{(k+1)/2} d\sigma = \min. \qquad (13)$$

Hence the $\infty^1$ related natural families forming a system $S_k$ on $\Sigma$, found by varying $h$, go over by conformal representation into $\infty^1$ natural families which are *not* usually related, that is, do *not* form a *whole* system $S_{k1}$ on $\Sigma_1$.

In the homothetic case, $\mu \equiv$ const., every whole system $S_k$ of $\infty^3$ curves on $\Sigma$ becomes a whole system $S_k$ (with the same value of $k$) of $\infty^3$ curves on $\Sigma_1$.

*In any non-homothic conformal representation of a surface* $\Sigma_1$ *upon a surface* $\Sigma$, *there is a unique conservative force whose whole system* $S_k$, *(where* $k \neq -1$ *and* $k \neq \infty$*), on* $\Sigma$ *is converted into a whole system* $S_k$ *(with the same value of* $k$*) of some conservative force on* $\Sigma_1$. *The work functions* $W$ *and* $W_1$ *on* $\Sigma$ *and* $\Sigma_1$ *are*

$$W = \mu^{2/(k+1)} = (d\sigma/ds)^{2/(k+1)}, \qquad W_1 = 1/W, \qquad (14)$$

*except for constant factors.*

There is no whole system $S_k$ on $\Sigma$ which by the non-homothetic conformal map becomes a whole system $S_{k_1}$ on $\Sigma_1$ with $k_1 \neq k$.

Similar remarks apply to velocity systems. *In any non-homothetic conformal representation of a surface* $\Sigma_1$ *upon a surface* $\Sigma$, *there is a unique conservative force whose whole velocity system* $S_\infty$ *on* $\Sigma$ *is converted into a whole velocity system* $S_\infty$ *of some conservative force on* $\Sigma_1$. *The work functions* $W$ *and* $W_1$ *on* $\Sigma$ *and* $\Sigma_1$ *are*

$$W = W_1 = \log \mu = \log d\sigma/ds, \qquad (15)$$

*except for constant factors.*

The work function $W$ of (15) in a conformal representation of [illegible] plane upon another plane is a harmonic function. Thus the fields of force obtained are inevitably Laplacian.

---

[*] Presented to the American Mathematical Society, 1949.

[1] Kasner, E., "Differential Geometric Aspects of Dynamics," *Amer. Math. Soc. Coll. Publ.*, 1913, 1934, 1948. See series of papers in *Trans. Amer. Math. Soc.*, **7–11**, 1906–1910.

[2] Kasner, E., "Physical Curves," PROC. NAT. ACAD. SCI., **33**, 246–251 (1947); and additional papers by Kasner and De Cicco in these PROCEEDINGS.

---

## ADDENDUM

### *Bounds for the Solutions of a Second-Order Linear Differential Equation*

It has been brought to the writer's attention that the first stated result of the note published under this title in these PROCEEDINGS, **35,** 190–191 (1949), was proved earlier by Z. Butlewski, see "Sur les intégrales d'une équation différentielle du second ordre," *Mathematica (Cluj)*, **12,** 42 (1936).

WALTER LEIGHTON

# NATIONAL ACADEMY OF SCIENCES ORGANIZATION

## July 1, 1949

### OFFICERS

| | *Term expires* |
|---|---|
| *President*—Alfred N. Richards | June 30, 1951 |
| *Vice President*—Edwin B. Wilson | June 30, 1953 |
| *Foreign Secretary*—Detlev W. Bronk | June 30, 1950 |
| *Home Secretary*—F. E. Wright | June 30, 1951 |
| *Treasurer*—William J. Robbins | June 30, 1952 |
| *Staff*—Executive Secretary—Raymund L. Zwemer | |
| Business Manager—G. D. Meid | |

### COUNCIL

| | | | |
|---|---|---|---|
| *Bronk, Detlev W. | (1950) | *Oppenheimer, J. Robert | (1951) |
| Goodpasture, Ernest W. | (1952) | *Richards, Alfred N. | (1951) |
| Hildebrand, Joel H. | (1952) | *Robbins, William J. | (1952) |
| Moore, Carl R. | (1951) | Van Slyke, Donald D. | (1950) |
| *Noyes, W. Albert, Jr. | (1950) | *Wilson, Edwin B. | (1953) |
| | | *Wright, F. E. | (1951) |

### MEMBERS

The number in parentheses, following year of election, indicates the Section to which the member belongs, as follows:

| | | | |
|---|---|---|---|
| (1) | Mathematics | (7) | Botany |
| (2) | Astronomy | (8) | Zoology and Anatomy |
| (3) | Physics | (9) | Physiology and Biochemistry |
| (4) | Engineering | (10) | Pathology and Bacteriology |
| (5) | Chemistry | (11) | Anthropology |
| (6) | Geology | (12) | Psychology |

**Abbot,** Charles Greeley, 1915 (2), 4604 Haring Lane, Bethesda, Md

**Adams,** Comfort Avery, 1930 (4), 417 West Price Street, Philadelphia 44, Pa.

**Adams,** Leason Heberling, 1943 (5), Geophysical Laboratory, Carnegie Institution of Washington, 2801 Upton Street, N. W., Washington 8, D. C.

**Adams, Roger,** 1929 (5), University of Illinois, Urbana, Ill.

---

* Members of the Executive Committee of the Council of the Academy.

**Adams, Walter** Sydney, 1917 (2), Mount Wilson Observatory, Pasadena 4, Calif.

**Addis, Thomas,** 1944 (10), Institute for Medical Research, Cedars of Lebanon Hospital, 4751 Fountain Avenue, Los Angeles 27, Calif.

**Adkins, Homer,** 1942 (5), 369 Chemistry Building, University of Wisconsin, Madison 6, Wis.

**Aitken, Robert** Grant, 1918 (2), 1109 Spruce Street, Berkeley 7, Calif.

**Albert,** Abraham **Adrian,** 1943 (1), Department of Mathematics, University of Chicago, Chicago 37, Ill.

**Alexander, James** Waddell, 1930 (1), 29 Cleveland Lane, Princeton, N. J.

**Allen,** Charles Elmer, 1924 (7), Room 103, Biology Building, University of Wisconsin, Madison 6, Wis.

**Allen, Eugene** Thomas, 1930 (6), The Irvington, 135 Pleasant Street, Arlington 74, Mass.

**Allison, Samuel King,** 1946 (3), Institute for Nuclear Studies, University of Chicago, Chicago 37, Ill.

**Alvarez, Luis** Walter, 1947 (3), Radiation Laboratory, University of California, Berkeley 4, Calif.

**Anderson, Carl** David, 1938 (3), California Institute of Technology, Pasadena 4, Calif.

**Anderson,** John August, 1928 (2), P. O. Box 332, Pasadena 17, Calif.

**Anderson, Rudolph** John, 1946 (9), Department of Chemistry, Yale University, New Haven, Conn.

**Armstrong, Charles,** 1944 (10), National Institute of Health, Bethesda 14, Md.

**Avery, Oswald** Theodore, 1933 (10), Hoods Hill Road, Nashville 5, Tenn.

**Babcock, Ernest** Brown, 1946 (7), Department of Genetics, University of California, Berkeley 4, Calif.

**Babcock, Harold** Delos, 1933 (2), Mount Wilson Observatory, Pasadena 4, Calif.

**Bacher, Robert** Fox, 1947 (3), Division of Physics, Mathematics, and Astrophysics, California Institute of Technology, Pasadena 4, Calif.

**Bachmann,** Werner Emmanuel, 1941 (5), University of Michigan, Ann Arbor, Mich.

**Bailey,** Irving Widmer, 1929 (7), Biological Laboratories, Harvard University, Cambridge 38, Mass.

**Bailey,** Liberty Hyde, 1917 (7), Bailey Hortorium, Ithaca, N. Y.

**Bainbridge, Kenneth** Tompkins, 1946 (3), Department of Physics, Harvard University, Cambridge 38, Mass.

**Ball,** Eric Glendinning, 1948 (9), Department of Biological Chemistry, Harvard University Medical School, Boston 15, Mass.

**Bancroft, Wilder** Dwight, 1920 (5), 7 East Avenue, Ithaca, N. Y.

**Bard, Philip,** 1944 (9), Johns Hopkins Medical School, 710 North Washington Street, Baltimore 5, Md.

**Bartelmez, George** William, 1949 (8), Department of Anatomy, University of Chicago, Chicago 37, Ill.

**Bartlett, Paul** Doughty, 1947 (5), Harvard University, 12 Oxford Street, Cambridge 38, Mass.

**Baxter, Gregory** Paul, 1916 (5), T. Jefferson Coolidge Jr. Memorial Laboratory, Harvard University, Cambridge 38, Mass.

**Beach, Frank** Ambrose, 1949 (12), Department of Psychology, Yale University, 333 Cedar Street, New Haven 11, Conn.

**Beadle,** George Wells, 1944 (7), Department of Biology, California Institute of Technology, Pasadena 4, Calif.

**Beams,** Jesse Wakefield, 1943 (3), Rouss Physical Laboratory, University of Virginia, Charlottesville, Va.

**Bell,** Eric Temple, 1927 (1), California Institute of Technology, Pasadena 4, Calif.

**Berkey, Charles** Peter, 1927 (6), Department of Geology, Columbia University, New York 27, N. Y.

**Berkner,** Lloyd Viel, 1948 (3), Carnegie Institution of Washington, Department of Terrestrial Magnetism, 5241 Broad Branch Road, N. W., Washington 15, D. C.

**Bethe,** Hans Albrecht, 1944 (3), Cornell University, Ithaca, N. Y.

**Bigelow, Henry** Bryant, 1931 (8) Museum of Comparative Zoology, Harvard University, Cambridge 38, Mass.

**Birge, Raymond** Thayer, 1932 (3), University of California, Berkeley 4, Calif.

**Bjerknes,** Jacob, 1947 (3), Department of Meteorology, University of California, Los Angeles 24, Calif.

**Blackwelder, Eliot,** 1936 (6), P. O. Box N, Stanford University, Calif.

**Blake, Francis** Gilman, 1947 (10), Yale University School of Medicine, 333 Cedar Street, New Haven 11, Conn.

**Blakeslee,** Albert Francis, 1929 (7), Genetics Experiment Station, Smith College, Northampton, Mass.

**Blalock, Alfred,** 1945 (10), Johns Hopkins Hospital, Baltimore 5, Md.

**Bliss, Gilbert** Ames, 1916 (1), Flossmoor, Ill.

**Bloch, Felix,** 1948 (3), Department of Physics, Stanford University, Stanford University, Calif.

**Bogert, Marston Taylor,** 1916 (5), 1158 Fifth Avenue, Apt. 14B, New York 29, N. Y.

**Bolton, Elmer K.,** 1946 (5), Department of Chemistry, E. I. du Pont de Nemours & Company, Wilmington, Del.

**Boring, Edwin** Garrigues, 1932 (12), Memorial Hall, Harvard University, Cambridge 38, Mass.

**Bowen,** Ira Sprague, 1936 (2), Mount Wilson Observatory, Pasadena 4, Calif.

**Bowen, Norman** Levi, 1935 (6), Geophysical Laboratory, Carnegie Institution of Washington, 2801 Upton Street, N. W., Washington 8, D. C.

**Bowman, Isaiah,** 1930 (6), Johns Hopkins University, Baltimore 18, Md.

**Bradley,** Wilmot Hyde, 1946 (6), United States Geological Survey, Washington 25, D. C.

**Breit,** Gregory, 1939 (3), Sloane Physics Laboratory, Yale University, New Haven, Conn.

**Bridgman,** Percy Williams, 1918 (3), Lyman Laboratory of Physics, Harvard University, Cambridge 38, Mass.

**Briggs,** Lyman James, 1942 (4), National Bureau of Standards, Washington 25, D. C.

**Brink,** Royal Alexander, 1947 (7), Department of Genetics, University of Wisconsin, Madison 6, Wis.

**Brode,** Robert Bigham, 1949 (3), Department of Physics, University of California, Berkeley 4, Calif.

**Bronk,** Detlev Wulf, 1939 (9), Johns Hopkins University, Baltimore 18, Md.

**Bucher,** Walter Hermann, 1938 (6), Department of Geology, Columbia University, New York 27, N. Y.

**Buckley,** Oliver Ellsworth, 1937 (4), Bell Telephone Laboratories, 463 West Street, New York 14, N. Y.

**Buddington,** Arthur Francis, 1943 (6), Princeton University, Princeton, N. J.

**Burkholder,** Paul Rufus, 1949 (7), Osborn Botanical Laboratory, Yale University, New Haven 11, Conn.

**Bush,** Vannevar, 1934 (4), Carnegie Institution of Washington, Washington 5, D. C.

**Byerly,** Perry, 1946 (6), Department of Seismology, University of California, Berkeley 4, Calif.

**Campbell,** Douglas Houghton, 1910 (7), Stanford University, Stanford University, Calif.

**Cannon,** Paul Roberts, 1946 (10), Department of Pathology, University of Chicago, Chicago 37, Ill.

**Carlson,** Anton Julius, 1920 (9), University of Chicago, Chicago 37, Ill.

**Carmichael,** Leonard, 1943 (12), Tufts College, Medford 55, Mass.

**Castle,** William Bosworth, 1939 (10), Boston City Hospital, Boston 18, Mass.

**Castle,** William Ernest, 1915 (8), 421 Spruce Street, Berkeley 8, Calif.

**Chandler,** William Henry, 1943 (7), College of Agriculture, University of California, 405 Hilgard Avenue, Los Angeles 24, Calif.

**Chaney,** Ralph Works, 1947 (6), Department of Paleontology, University of California, Berkeley 4, Calif.

**Child,** Charles Manning, 1935 (8), Jordan Hall, Stanford University, Stanford University, Calif.

**Clark,** William Mansfield, 1928 (9), Johns Hopkins Medical School, 710 North Washington Street, Baltimore 5, Md.

**Clarke,** Hans Thacher, 1942 (9), College of Physicians and Surgeons, 630 West 168th Street, New York 32, N. Y.

**Cleland,** Ralph Erskine, 1942 (7), Indiana University, Bloomington, Ind.

**Coble,** Arthur Byron, 1924 (1), Haverford College, Haverford, Pa.

**Coblentz,** William Weber, 1930 (3), 2737 Macomb Street, N. W., Washington 8, D. C.

**Cochrane,** Edward Lull, 1945 (4), Department of Naval Architecture and Marine Engineering, Massachusetts Institute of Technology, Cambridge 39, Mass.

**Coggeshall,** Lowell Thelwell, 1949 (10), Division of Biological Sciences, University of Chicago, Chicago 37, Ill.

**Cohn, Edwin** Joseph, 1943 (9), Harvard Medical School, Boston 15, Mass.

**Cole, Rufus,** 1922 (10), Mt. Kisco, N. Y.

**Compton, Arthur** Holly, 1927 (3), Washington University, St. Louis 5, Mo.

**Compton, Karl** Taylor, 1924 (3), Research & Development Board, Room 3E600, Pentagon, Washington 25, D. C.

**Conant, James** Bryant, 1929 (5), Harvard University, Cambridge 38, Mass.

**Condon,** Edward Uhler, 1944 (3), National Bureau of Standards, Washington 25, D. C.

**Conklin, Edwin** Grant, 1908 (8), Princeton University, Princeton, N. J.

**Coolidge,** William David, 1925 (3), 1480 Lenox Road, Schenectady, N. Y.

**Cope, Arthur** Clay, 1947 (5), Department of Chemistry, Massachusetts Institute of Technology, Cambridge 39, Mass.

**Cori, Carl** Ferdinand, 1940 (9), School of Medicine, Washington University, Euclid Avenue and Kingshighway, St. Louis 10, Mo.

**Cori, Gerty** Theresa, 1948 (9), Department of Biological Chemistry, School of Medicine, Washington University, Euclid Avenue and Kingshighway, St. Louis 10, Mo.

**Corner, George** Washington, 1940 (8), Department of Embryology, Carnegie Institution of Washington, Wolfe and Madison Streets, Baltimore 5, Md.

**Couch, John** Nathaniel, 1943 (7), University of North Carolina, Chapel Hill, N. C.

**Crew, Henry,** 1909 (3), 620 Library Place, Evanston, Ill.

**Curme, George** Oliver, **Jr.,** 1944 (4), Carbide & Carbon Chemical Corporation, 30 East 42nd Street, New York 11, N. Y.

**Daly, Reginald** Aldworth, 1925 (6), 23 Hawthorn Street, Cambridge 38, Mass.

**Danforth,** Charles Haskell, 1942 (8), Department of Anatomy, Stanford University, Stanford University, Calif.

**Daniels, Farrington,** 1947 (5), Department of Chemistry, University of Wisconsin, Madison 6, Wis.

**Davis, Bergen,** 1929 (3), 44 Morningside Drive, New York 25, N. Y.

**Davis, Hallowell,** 1948 (9), Central Institute for the Deaf, 818 South Kingshighway, St. Louis 10, Mo.

**Davisson,** Clinton Joseph, 1929 (3), Department of Physics, University of Virginia, Charlottesville, Va.

**Day, Arthur** Louis, 1911 (6), 9113 Old Georgetown Road, Bethesda 14, Md.

**Debye, Peter,** 1947* (5), Baker Laboratory, Cornell University, Ithaca, N. Y.

**Delbrück, Max,** 1949 (7), Kerckhoff Laboratories of Biology, California Institute of Technology, Pasadena 4, Calif.

---

* Elected a foreign associate in 1931; became member on naturalizatior.

**Demerec, Milislav,** 1946 (8), Department of Genetics, Carnegie Institution of Washington, Cold Spring Harbor, N. Y.

**Dempster, Arthur** Jeffrey, 1937 (3), University of Chicago, Chicago 37, Ill.

**Detwiler,** Samuel Randall, 1932 (8), College of Physicians and Surgeons, 630 West 168th Street, New York 32, N. Y.

**Dobzhansky, Theodosius,** 1943 (8), Department of Zoology, Columbia University, New York 27, N. Y.

**Dochez,** Alphonse Raymond, 1933 (10), Columbia University, Presbyterian Hospital, 620 West 168th Street, New York 32, N. Y.

**Dodge,** Bernard Ogilvie, 1933 (7), New York Botanical Garden, Bronx Park (Fordham Station), New York 58, N. Y.

**Doisy,** Edward Adelbert, 1938 (9), St. Louis University School of Medicine, 1402 South Grand Boulevard, St. Louis 4, Mo.

**Douglas, Jesse,** 1946 (1), 2610 Glenwood Road, Brooklyn 10, N. Y.

**Dryden, Hugh** Latimer, 1944 (4), National Advisory Committee for Aeronautics, 1724 F Street, N. W., Washington 25, D. C.

**DuBois, Eugene** Floyd, 1933 (9), Cornell University Medical College, 1300 York Avenue, New York 21, N. Y.

**Dubos, René** Jules, 1941 (10), Rockefeller Institute for Medical Research, 66th Street and York Avenue, New York 21, N. Y.

**DuBridge,** Lee Alvin, 1943 (3), California Institute of Technology, Pasadena 4, Calif.

**Duggar,** Benjamin Minge, 1927 (7), Lederle Laboratories, Incorporated, Pearl River, N. Y.

**Dunbar, Carl** Owen, 1944 (6), Peabody Museum, Yale University, New Haven, Conn.

**Dunn, Gano,** 1919 (4), 80 Broad Street, New York 4, N. Y.

**Dunn,** Leslie Clarence, 1943 (8), Columbia University, New York 27, N. Y.

**Dunning,** John Ray, 1948 (3), Department of Physics, Columbia University, New York 27, N. Y.

**Durand,** William Frederick, 1917 (4), Stanford University, Stanford University, Calif.

**du Vigneaud, Vincent,** 1944 (9), Cornell University Medical College, New York 21, N. Y.

**Einstein, Albert,** 1942* (3), The Institute for Advanced Study, Princeton, N. J.

**Eisenhart,** Luther Pfahler, 1922 (1), 25 Alexander Street, Princeton, N. J.

**Elderfield, Robert** Cooley, 1949 (5), Department of Chemistry, Columbia University, New York 27, N. Y.

**Elvehjem,** Conrad Arnold, 1942 (9), Department of Biochemistry, University of Wisconsin, Madison 6, Wis.

**Epstein, Paul** Sophus, 1930 (3), 1484 Oakdale Street, Pasadena 4, Calif.

**Erlanger, Joseph,** 1922 (9), Washington University School of Medicine, 4580 Scott Avenue, St. Louis 10, Mo.

---

* Elected a foreign associate in 1922; became member on naturalization.

**Evans, Griffith** Conrad, 1933 (1), Department of Mathematics, University of California, Berkeley 4, Calif.

**Evans, Herbert** McLean, 1927 (9), Institute of Experimental Biology, University of California, Berkeley 4, Calif.

**Ewing,** William **Maurice,** 1948 (6), Department of Geology, Columbia University, New York 27, N. Y.

**Eyring, Henry,** 1945 (5), Graduate School, University of Utah, Salt Lake City, Utah

**Fenn, Wallace** Osgood, 1943 (9), School of Medicine & Dentistry, University of Rochester, 260 Crittenden Boulevard, Rochester 7, N. Y.

**Fermi, Enrico,** 1945 (3), Institute for Nuclear Studies, University of Chicago, Chicago 37, Ill.

**Fieser, Louis** Frederick, 1940 (5), Harvard University, Cambridge 38, Mass.

**Fleming, John** Adam, 1938 (2), Carnegie Institution of Washington, 1530 P Street, N. W., Washington 5, D. C.

**Fletcher, Harvey,** 1935 (4), Bell Telephone Laboratories, Murray Hill, N. J.

**Folkers, Karl** August, 1948 (5), Research & Development Division, Merck and Company, Incorporated, Rahway, N. J.

**Foote, Paul** Darwin, 1943 (4), Gulf Research & Development Company, P. O. Drawer, 2038, Pittsburgh 30, Pa.

**Forbes, Alexander,** 1936 (9), Harland Street, Milton, Mass.

**Francis, Thomas, Jr.,** 1948 (10), Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Mich.

**Franck, James,** 1944 (3), Department of Chemistry, University of Chicago, Chicago 37, Ill.

**Fred,** Edwin Broun, 1931 (7), University of Wisconsin, Madison 6, Wis.

**Fuson,** Reynold Clayton, 1944 (5), 263 Noyes Laboratory, University of Illinois, Urbana, Ill.

**Gamble, James** Lawder, 1945 (10), 33 Edge Hill Road, Brookline, Mass.

**Gasser, Herbert** Spencer, 1934 (9), Rockefeller Institute for Medical Research, 66th Street and York Avenue, New York 21, N. Y.

**Gesell, Arnold,** 1947 (12), Clinic of Child Development, Yale University School of Medicine, 14 Davenport Avenue, New Haven, Conn.

**Giauque, William** Francis, 1936 (5), University of California, Berkeley 4, Calif.

**Gibbs, William Francis,** 1949 (4), One Broadway, New York, N. Y.

**Gilliland, Edwin** Richard, 1948 (4), Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge 39, Mass.

**Gilluly, James,** 1947 (6), Department of Geology, University of California, Los Angeles 24, Calif.

**Gilman, Henry,** 1945 (5), Department of Chemistry, Iowa State College, Ames, Iowa

**Goldschmidt, Richard** Benedikt, 1947 (8), Department of Zoology, University of California, Berkeley 4, Calif.

**Goodpasture, Ernest** William, 1937 (10), Vanderbilt University, Nashville 4, Tenn.

**Goudsmit, Samuel** Abraham, 1947 (3), Theoretical Physics Division, Brookhaven National Laboratory, Upton, Long Island, N. Y.

**Graham,** Clarence Henry, 1946 (12), Department of Psychology, Columbia University, New York 27, N. Y.

**Graham, Evarts** Ambrose, 1941 (10), Barnes Hospital, 600 South Kingshighway, St. Louis 10, Mo.

**Gregory, William King,** 1927 (6), American Museum of Natural History, 77th Street and Central Park West, New York 24, N. Y.

**Gutenberg, Beno,** 1945 (6), Seismological Laboratory, 220 North San Rafael Avenue, Pasadena 2, Calif.

**Hammett, Louis Plack,** 1943 (5), Department of Chemistry, Columbia University, New York 27, N. Y.

**Hansen, William** Webster, 1949* (3), Department of Physics, Stanford University, Stanford, Calif.

**Harkins,** William Draper, 1921 (5), University of Chicago, Chicago 37, Ill.

**Harrison, Ross** Granville, 1913 (8), Osborn Zoological Laboratory, Yale University, New Haven, Conn.

**Hart, Edwin** Bret, 1944 (9), Biochemistry Building, University of Wisconsin, Madison 6, Wis.

**Hartline,** Haldan **Keffer,** 1948 (9), Department of Biophysics, Biological Laboratories, Johns Hopkins University, Baltimore 18, Md.

**Hartman, Carl** Gottfried, 1937 (8), Ortho Research Foundation, Raritan, N. J.

**Harvey,** Edmund **Newton,** 1934 (8), Guyot Hall, Princeton University, Princeton, N. J.

**Hastings, Albert Baird,** 1939 (9), Department of Biological Chemistry, Harvard University Medical School, Boston 15, Mass.

**Heidelberger, Michael,** 1942 (10), College of Physicians and Surgeons, 620 West 168th Street, New York 32, N. Y.

**Herrick,** Charles **Judson,** 1918 (8), 236 Morningside Drive, S. E., Grand Rapids, Mich.

**Herty, Charles Holmes, Jr.,** 1947 (4), Bethlehem Steel Company, Incorporated, Bethlehem, Pa.

**Hewett,** Donnel Foster, 1937 (6), 1460 Rose Villa Street, Pasadena 4, Calif.

**Hildebrand, Joel** Henry, 1929 (5), Gilman Hall, University of California, Berkeley 4, Calif.

**Hilgard,** Ernest Ropiequet, 1948 (12), Department of Psychology, Stanford University, Stanford University, Calif.

**Hisaw, Frederick** Lee, 1947 (8), Biological Laboratories, Harvard University, 16 Divinity Avenue, Cambridge 38, Mass.

**Hoagland,** Dennis Robert, 1934 (7), Division of Plant Nutrition, University of California, Berkeley 4, Calif.

---

* Died May 23, 1949.

**Hooton, Earnest** Albert, 1935 (11), Peabody Museum, Harvard University, Cambridge 38, Mass.

**Hoover, Herbert** Clark, 1922 (4), Waldorf Astoria Towers, New York, N. Y.

**Horsfall, Frank** Lappin, **Jr.,** 1948 (10), Hospital of the Rockefeller Institute for Medical Research, 66th Street and York Avenue, New York 21, N. Y.

**Houston, William** Vermillion, 1943 (3), The Rice Institute, Houston, Tex.

**Hovgaard, William,** 1929 (4), Countryside, Summit, N. J.

**Hubble, Edwin** Powell, 1927 (2), Mount Wilson Observatory, Pasadena, 4, Calif.

**Hudson, Claude** Silbert, 1927 (5), National Institute of Health, Bethesda 14, Md.

**Huggins, Charles** Brenton, 1949 (10), Department of Surgery, University of Chicago, 950 East 59th Street, Chicago 37, Ill.

**Hulett, George** Augustus, 1922 (5), 44 Washington Road, Princeton, N. J.

**Hull, Albert** Wallace, 1929 (3), Research Laboratory, General Electric Company, Schenectady 5, N. Y.

**Hull, Clark** Leonard, 1936 (12), Institute of Human Relations, Yale University, 333 Cedar Street, New Haven, Conn.

**Hunsaker, Jerome** Clark, 1935 (4), Massachusetts Institute of Technology, Cambridge 39, Mass.

**Hunter, Walter** Samuel, 1935 (12), Brown University, Providence 12, R. I.

**Ipatieff,** Vladimir Nikolaevich, 1939 (5), Universal Oil Products Company, Riverside, Ill.

**Ives, Herbert** Eugene, 1933 (3), 32 Laurel Place, Upper Montclair, N. J.

**Jacobs,** Merkel Henry, 1939 (8), School of Medicine, University of Pennsylvania, Philadelphia 4, Pa.

**Jacobs, Walter** Abraham, 1932 (5), Rockefeller Institute for Medical Research, 66th Street and York Avenue, New York 21, N. Y.

**Jeffries, Zay,** 1939 (4), General Electric Company, 1 Plastics Avenue, Pittsfield, Mass.

**Jewett, Frank** Baldwin, 1918 (4), 388 Hobart Avenue, Short Hills, N. J.

**Johnson, John** Raven, 1948 (5), Department of Chemistry, Cornell University, Ithaca, N. Y.

**Jones,** Donald Forsha, 1939 (7), Box 1106, Department of Genetics, Connecticut Agricultural Experiment Station, New Haven 4, Conn.

**Joy, Alfred** Harrison, 1944 (2), Mount Wilson Observatory, Pasadena 4, Calif.

**Kasner, Edward,** 1917 (1), 430 West 116th Street, New York 27, N. Y.

**Kelley,** Walter Pearson, 1943 (6), 120 Hilgard Hall, University of California, Berkeley 4, Calif.

**Kelly, Mervin** J., 1945 (4), Bell Telephone Laboratories, 463 West Street, New York 14, N. Y.

**Kelser, Raymond** Alexander, 1948 (10), School of Veterinary Medicine, University of Pennsylvania, Philadelphia 4, Pa.

**Kemble,** Edwin Crawford, 1931 (3), Physics Laboratories, Harvard University, Cambridge 38, Mass.

**Kettering,** Charles Franklin, 1928 (4), General Motors Corporation, Detroit 2, Mich.

**Keyes,** Frederick George, 1930 (5), Massachusetts Institute of Technology, Cambridge 39, Mass.

**Kharasch,** Morris Selig, 1946 (5), Department of Chemistry, University of Chicago, Chicago 37, Ill.

**Kidder,** Alfred Vincent, 1936 (11), 10 Frisbie Place, Cambridge 38, Mass.

**King, Arthur** Scott, 1941 (2), Mount Wilson Observatory, Pasadena 4, Calif.

**Kirkwood, John** Gamble, 1942 (5), Division of Chemistry, California Institute of Technology, Pasadena 4, Calif.

**Kistiakowsky,** George Bogdan, 1939 (5), Department of Chemistry, 12 Oxford Street, Harvard University, Cambridge 38, Mass.

**Knopf, Adolph,** 1931 (6), Yale University, New Haven, Conn.

**Köhler, Wolfgang,** 1947 (12), Swarthmore College, Swarthmore, Pa.

**Kraus, Charles** August, 1925 (5), Brown University, Providence 12, R. I.

**Kroeber,** Alfred L., 1928 (11), Department of Anthropology, Columbia University, New York 27, N. Y.

**Kunkel,** Louis Otto, 1932, (7), Rockefeller Institute for Medical Research, 66th Street and York Avenue, New York 21, N. Y.

**Lamb, Arthur** Becket, 1924 (5), Chemical Laboratory, Harvard University, Cambridge 38, Mass.

**Lambert, Walter** Davis, 1949 (2), P. O. Box 687, Canaan, Conn.

**LaMer,** Victor Kuhn, 1945 (5), Department of Chemistry, Columbia University, New York 27, N. Y.

**Langmuir, Irving,** 1918 (5), General Electric Company, Schenectady 5, N. Y.

**Larsen, Esper S., Jr.,** 1944 (6), Geological Museum, Oxford Street, Cambridge 38, Mass.

**Lashley,** Karl Spencer, 1930 (12), Yerkes Laboratories of Primate Biology, Orange Park, Fla.

**Latimer, Wendell** Mitchell, 1940 (5), University of California, Berkeley 4, Calif.

**Lauritsen, Charles** Christian, 1941 (3), California Institute of Technology, Pasadena 4, Calif.

**Lawrence, Ernest** Orlando, 1934 (3), Radiation Laboratory, University of California, Berkeley 4, Calif.

**Lawson, Andrew** Cowper, 1924 (6), University of California, Berkeley 4, Calif.

**Lefschetz, Solomon,** 1925 (1), Fine Hall, 129 Broadmead Street, Princeton, N. J.

**Leith,** Charles **Kenneth,** 1920 (6), Wardman Park Hotel, Washington 8, D. C.

**Leuschner, Armin Otto,** 1913 (2), 1816 Scenic Avenue, Berkeley 9, Calif.

**Lewis, Howard** Bishop, 1949 (9), Department of Biological Chemistry, Medical School, University of Michigan, Ann Arbor, Mich.

**Lewis, Warren** Harmon, 1936 (8), The Wistar Institute of Anatomy and Biology, Philadelphia 4, Pa.

**Lewis,** Warren Kendall, 1938 (4), Massachusetts Institute of Technology, Cambridge 39, Mass.

**Lind,** Samuel Colville, 1930 (5), P. O. Box P, Oak Ridge, Tenn.

**Link, Karl Paul,** 1946 (9), Department of Biochemistry, Agricultural Experiment Station, University of Wisconsin, Madison, Wis.

**Linton, Ralph,** 1945 (11), Institute of Human Relations, 333 Cedar Street, New Haven 11, Conn.

**Little,** Clarence Cook, 1945 (10), Roscoe B. Jackson Memorial Laboratory, Bar Harbor, Me.

**Loeb, Leo,** 1937 (10), 40 Crestwood Drive, St. Louis 5, Mo.

**Loeb, Robert** Frederick, 1946 (9), College of Physicians and Surgeons, 620 West 168th Street, New York 32, N. Y.

**Long,** Cyril **Nor**man Hugh, 1948 (9), Yale University School of Medicine, 333 Cedar Street, New Haven 11, Conn.

**Long, Esmond** Ray, 1946 (10), Henry Phipps Institute, 7th and Lombard Streets, Philadelphia 47, Pa.

**Longcope, Warfield** Theobald, 1943 (10), Cornhill Farm, Lee, Mass.

**Longsworth, Lewis** Gibson, 1947 (5), Rockefeller Institute for Medical Research, 66th Street and York Avenue, New York 21, N. Y.

**Longwell, Chester** Ray, 1935 (6), Kirtland Hall, Yale University, New Haven, Conn.

**Loomis, Alfred** Lee, 1941 (4), The Loomis Laboratory, Room 2420, 14 Wall St., New York 5, N. Y.

**Loomis,** Francis Wheeler, 1949 (3), Department of Physics, University of Illinois, Urbana, Ill.

**Lovering, Thomas** Seward, 1949 (6), 4065 University Station, Tuscon, Ariz.

**Lowie, Robert** Harry, 1931 (11), University of California, Berkeley 4, Calif.

**Lyman, Theodore,** 1917 (3), Research Laboratory of Physics, Harvard University, Cambridge 38, Mass.

**McClintock, Barbara,** 1944 (7), Carnegie Institution, Cold Spring Harbor, Long Island, N. Y.

**McCollum,** Elmer Verner, 1920 (9), Gilman Hall, Johns Hopkins University, Baltimore 18, Md.

**McElvain,** Samuel **M**arion, 1949 (5), Department of Chemistry, University of Wisconsin, Madison 6, Wis.

**McMillan, Edwin** Mattison, 1947 (3), Radiation Laboratory, University of California, Berkeley 4, Calif.

**McShane,** Edward James, 1948 (1), School of Mathematics, University of Virginia, Charlottesville, Va.

**MacInnes, Duncan Arthur,** 1937 (5), Rockefeller Institute for Medical Research, 66th Street and York Avenue, New York 21, N. Y.

**Mac Lane, Saunders,** 1949 (1), Department of Mathematics, University of Chicago, Chicago 37, Ill.

**MacNider, William deBerniere,** 1938 (9), University of North Carolina, Chapel Hill, N. C.

**Macelwane, James** Bernard, **S.J.,** 1944 (6), 221 North Grand Boulevard, St. Louis 3, Mo.

**Mangelsdorf, Paul** Christoph, 1945 (7), Botanical Museum, Harvard University, Cambridge 38, Mass.

**Marshall, Eli Kennerly, Jr.,** 1943 (9), School of Medicine, Johns Hopkins University, 710 North Washington Street, Baltimore 5, Md.

**Marvel,** Carl **Shipp,** 1938 (5), Department of Chemistry, University of Illinois, Urbana, Ill.

**Mason, Max,** 1923 (3), California Institute of Technology, Pasadena 4, Calif.

**Mayall, Nicholas Ulrich,** 1949 (2), University of California, Lick Observatory, Mount Hamilton, Calif.

**Mayer, Joseph** Edward, 1946 (5), Department of Chemistry, University of Chicago, Chicago 37, Ill.

**Maynard, Leonard** Amby, 1944 (9), School of Nutrition, Cornell University, Ithaca, N. Y.

**Mead, Warren** Judson, 1939 (6), Massachusetts Institute of Technology, Cambridge 39, Mass.

**Meek, Walter** Joseph, 1947 (9), Department of Physiology, University of Wisconsin, Madison 6, Wis.

**Mendenhall,** Walter Curran, 1932 (6), 9 East Lenox Street, Chevy Chase 15, Md.

**Menzel, Donald** Howard, 1948 (2), Harvard College Observatory, Cambridge 38, Mass.

**Merica, Paul** Dyer, 1942 (4), 67 Wall Street, New York 5, N. Y.

**Merrill,** Elmer Drew, 1923 (7), Arnold Arboretum, Jamaica Plain, Mass.

**Merrill, Paul** Willard, 1929 (2), Mount Wilson Observatory, Pasadena 4, Calif.

**Metz,** Charles William, 1948 (8), University of Pennsylvania, Zoological Laboratory, 38th Street and Woodland Avenue, Philadelphia 4, Pa.

**Meyer, Karl** Friederich, 1940 (10), George Williams Hooper Foundation, University of California Medical Center, San Francisco 22, Calif.

**Meyerhof, Otto,** 1949 (9), Department of Physiological Chemistry, The School of Medicine, University of Pennsylvania, Philadelphia 4, Pa.

**Michaelis, Leonor,** 1943 (9), Rockefeller Institute for Medical Research, 66th Street and York Avenue, New York 21, N. Y.

**Miles, Walter** Richard, 1933 (12), Yale University School of Medicine, 333 Cedar Street, New Haven 11, Conn.

**Miller,** George Abram, 1921 (1), 1203 West Illinois Street, Urbana, Ill.

**Millikan,** Robert Andrews, 1915 (3), California Institute of Technology, Pasadena 4, Calif.

**Minot, George** Richards, 1937 (10), Thorndike Memorial Laboratory, Boston City Hospital, Boston 18, Mass.

**Mitchell,** Samuel Alfred, 1933 (2), Leander McCormick Observatory, University Station, Charlottesville, Va.

**Moore,** Carl Richard, 1944 (8), University of Chicago, Chicago 37, Ill.

**Moore,** Robert Lee, 1931 (1), University of Texas, Austin 12, Tex.

**Morse,** Harold Marston, 1932 (1), The Institute for Advanced Study, Princeton, N. J.

**Moulton,** Forest Ray, 1910 (2), 3027 N Street, N. W., Washington 7, D. C.

**Mueller,** John Howard, 1945 (10), Harvard Medical School, 25 Shattuck Street, Boston 15, Mass.

**Muller,** Hermann Joseph, 1931 (8), Science Hall 101, Indiana University, Bloomington, Ind.

**Mulliken,** Robert Sanderson, 1936 (3), University of Chicago, Chicago 37, Ill.

**Murnaghan,** Francis Dominic, 1942 (1), Johns Hopkins University, Baltimore 18, Md.

**Murphy,** James Bumgardner, 1940 (10), Rockefeller Institute for Medical Research, 66th Street and York Avenue, New York 21, N. Y.

**Nicholas,** John Spangler, 1949 (8), Osborn Zoological Laboratory, Yale University, New Haven 11, Conn.

**Nicholson,** Seth Barnes, 1937 (2), Mount Wilson Observatory, Pasadena 4, Calif.

**Northrop,** John Howard, 1934 (9), Rockefeller Institute for Medical Research, Princeton, N. J.

**Novy,** Frederick George, 1924 (10), 721 Forest Avenue, Ann Arbor, Mich.

**Noyes,** William Albert, Jr., 1943 (5), Department of Chemistry, University of Rochester, Rochester 3, N. Y.

**Oncley,** John Lawrence, 1947 (9), Department of Physical Chemistry, Harvard Medical School, Boston 15, Mass.

**Onsager,** Lars, 1947 (5), Sterling Chemistry Laboratory, Yale University, 225 Prospect Street, New Haven, Conn.

**Opie,** Eugene Lindsay, 1923 (10), Rockefeller Institute for Medical Research, 66th Street and York Avenue, New York 21, N. Y.

**Oppenheimer,** J. Robert, 1941 (3), The Institute for Advanced Study, Princeton, N. J.

**Osterhout,** Winthrop John Vanleuven, 1919 (7), Rockefeller Institute for Medical Research, 66th Street and York Avenue, New York 21, N. Y.

**Painter,** Theophilus Shickel, 1938 (8), University of Texas, Austin 12, Tex.

**Palache,** Charles, 1934 (6), Harvard University, Cambridge 38, Mass.

**Parker,** George Howard, 1913 (8), Harvard Biological Laboratories, Divinity Avenue, Cambridge 38, Mass.

**Patterson,** John Thomas, 1941 (8), Department of Zoology, University of Texas, Austin 12, Tex.

**Paul,** John Rodman, 1945 (10), Yale University School of Medicine 333 Cedar Street, New Haven 11, Conn.

**Pauling, Linus,** 1933 (5), Crellin Laboratory, California Institute of Technology, Pasadena 4, Calif.

**Pegram, George** Braxton, 1949 (3), Columbia University, New York 27, N. Y.

**Peters, John** Punnett, 1947 (9), Department of Internal Medicine, Yale University School of Medicine, New Haven 11, Conn.

**Pierce,** George Washington, 1920 (3), Cruft Laboratory, Harvard University, Cambridge 38, Mass.

**Piggot, Charles** Snowden, 1946 (3), Committee on Geophysical Sciences, Research and Development Board, The Pentagon, Room 3D571, Washington 25, D. C.

**Pillsbury,** Walter Bowers, 1925 (12), University of Michigan, Ann Arbor, Mich.

**Pitzer,** Kenneth, Sanborn, 1949 (5), Division of Research, U. S. Atomic Energy Commission, 1901 Constitution Avenue, Washington 25, D. C.

**Rabi,** Isidor Isaac, 1940 (3), Department of Physics, Columbia University, New York 27, N. Y.

**Raper, Kenneth** Bryan, 1949 (7), Northern Regional Research Laboratory, U. S. Department of Agriculture, Peoria 5, Ill.

**Reeside, John** Bernard, **Jr.,** 1945 (6), Box 177, Hyattsville, Md.

**Reichelderfer, Francis** Wilton, 1945 (3), Room 310, U. S. Weather Bureau, Washington 25, D. C.

**Rhoades,** Marcus Morton, 1946 (7), Department of Botany, University of Illinois, Urbana, Ill.

**Richards, Alfred** Newton, 1927 (9), School of Medicine, University of Pennsylvania, Philadelphia 4, Pa.

**Richter, Curt** Paul, 1948 (12), The Johns Hopkins Hospital, Baltimore 5, Md.

**Riddle, Oscar,** 1939 (8), Route 4, Plant City, Fla.

**Ritt,** Joseph Fels, 1933 (1), Columbia University, New York 27, N. Y.

**Rivers, Thomas** Milton, 1934 (10), Rockefeller Institute for Medical Research, 66th Street and York Avenue, New York 21, N. Y.

**Robbins, William** Jacob, 1940 (7), New York Botanical Garden, Bronx Park (Fordham Station), New York 58, N. Y.

**Robertson,** Oswald Hope, 1943 (10), Department of Medicine, University of Chicago, Chicago 37, Ill.

**Rodebush,** Worth Huff, 1938 (5), University of Illinois, Urbana, Ill.

**Romer, Alfred** Sherwood, 1944 (8), Museum of Comparative Zoology, Oxford Street, Cambridge 38, Mass.

**Rose, William** Cumming, 1936 (9), University of Illinois, Urbana, Ill.

**Ross, Frank** Elmore, 1930 (2), Mount Wilson Observatory, Pasadena 4, Calif.

**Rossby,** Carl-Gustaf Arvid, 1943 (3), Institute of Meteorology, University of Chicago, Chicago 37, Ill.

**Rous,** Francis **Peyton,** 1927 (10), Rockefeller Institute for Medical Research, 66th Street and York Avenue, New York 21, N. Y.

**Rubey, William** Walden, 1945 (6), U. S. Geological Survey, Washington 25, D. C.

**Ruedemann, Rudolf,** 1928 (6), 161 Dana Avenue, Albany 3, N. Y.

**Russell, Henry Norris,** 1918 (2), Princeton University, Princeton, N. J.

**Sabin, Florence** Rena, 1925 (10), 1333 East 10th Avenue, Denver 3, Colo.

**Saunders, Frederick** Albert, 1925 (3), South Hadley, Mass.

**Savage,** John Lucian, 1949 (4), 1651 Dahlia Street, Denver 7, Colo.

**Sax, Karl,** 1941 (7), Harvard University, Cambridge 38, Mass.

**Scatchard, George,** 1946 (5), Department of Chemistry, Massachusetts Institute of Technology, Cambridge 39, Mass.

**Schlesinger,** Hermann Irving, 1948 (5), Department of Chemistry, University of Chicago, Chicago 37, Ill.

**Schmidt, Carl** Frederic, 1949 (9), Laboratory of Pharmacology, School of Medicine, University of Pennsylvania, Philadelphia 4, Pa.

**Schmitt, Francis** Otto, 1948 (8), Department of Biology, Massachusetts Institute of Technology, Cambridge 39, Mass.

**Schultz, Adolph** Hans, 1939 (11), Johns Hopkins Medical School, Baltimore 5, Md.

**Schwinger, Julian,** 1949 (3), Department of Physics, Harvard University, Cambridge 38, Mass.

**Seaborg, Glenn** Theodore, 1948 (5), Department of Chemistry, University of California, Berkeley 4, Calif.

**Seares,** Frederick Hanley, 1919 (2), 351 Palmetto Drive, Pasadena 2, Calif.

**Seashore,** Carl Emil, 1922 (12), State University of Iowa, Iowa City, Iowa

**Shaffer, Philip** Anderson, 1928 (9), Washington University Medical School, St. Louis 10, Mo.

**Shapiro,** Harry Lionel, 1949 (11), Department of Anthropology, American Museum of Natural History, Central Park West at 79th Street, New York 24, N. Y.

**Shapley, Harlow,** 1924 (2), Harvard College Observatory, Cambridge 38, Mass.

**Sherman,** Henry Clapp, 1933 (9), Columbia University, New York 27, N. Y.

**Shope, Richard** Edwin, 1940 (10), Box 323, Kingston, N. J.

**Simpson,** George Gaylord, 1941 (6), American Museum of Natural History, 77th Street and Central Park West, New York 24, N. Y.

**Sinnott, Edmund Ware,** 1936 (7), Osborn Botanical Laboratory, Yale University, New Haven, Conn.

**Slater, John** Clarke, 1932 (3), Massachusetts Institute of Technology, Cambridge 39, Mass.

**Slepian,** Joseph, 1941 (4), Westinghouse Electric Corporation, East Pittsburgh, Pa.

**Slichter, Louis** Byrne, 1944 (3), 1446 North Amalfi Drive, Pacific Palisades, Calif.

**Slipher,** Vesto **Melvin,** 1921 (2), Lowell Observatory, Flagstaff, Ariz.

**Small, Lyndon** Frederick, 1941 (5), National Institutes of Health, Bethesda 14, Md.

**Smith, Gilbert Morgan**, 1948 (7), School of Biological Sciences, Stanford University, Stanford University, Calif.

**Smith, Homer** William, 1945 (9), 477 First Avenue, New York 16, N. Y.

**Smith, Lee Irvin**, 1944 (5), School of Chemistry, University of Minnesota, Minneapolis 14, Minn.

**Smith, Paul Althaus**, 1947 (1) Department of Mathematics, Columbia University, New York 27, N. Y.

**Smith, Philip** Edward, 1939 (8), College of Physicians and Surgeons, 630 West 168th Street, New York 32, N. Y.

**Soderberg,** Carl **Richard,** 1947 (4), Massachusetts Institute of Technology, Cambridge 39, Mass.

**Sonneborn, Tracy** Morton, 1946 (8), Department of Zoology, Indiana University, Bloomington, Ind.

**Spier, Leslie,** 1946 (11), P. O. Box 880, Santa Cruz, Calif.

**Stadie, William** Christopher, 1945 (9), 821 Maloney Clinic, 36th and Spruce Streets, Philadelphia 4, Pa.

**Stadler,** Lewis John, 1938 (7), University of Missouri, Columbia, Mo.

**Stakman,** Elvin Charles, 1934 (7), University Farm, St. Paul 8, Minn.

**Stanley, Wendell** Meredith, 1941 (9), Department of Biochemistry, University of California, Berkeley 4, Calif.

**Stebbins, Joel,** 1920 (2), Lick Observatory, Mt. Hamilton, Calif.

**Stern, Curt,** 1948 (8), Department of Zoology, University of California, Berkeley 4, Calif.

**Stern, Otto,** 1945 (3), 759 Cragmont Avenue, Berkeley, Calif.

**Stevens,** Stanley **Smith,** 1946 (12), Department of Psychology, Harvard University, Cambridge 38, Mass.

**Stewart,** George **Walter,** 1938 (3), State University of Iowa, Iowa City, Iowa

**Stock, Chester,** 1948 (6), Division of Geological Sciences, California Institute of Technology, Pasadena 4, Calif.

**Stone, Calvin** Perry, 1943 (12), Stanford University, Stanford University, Calif.

**Stone, Marshall** Harvey, 1938 (1), 313 Eckhart Hall, University of Chicago, Chicago 37, Ill.

**Struve, Otto,** 1937 (2), Yerkes Observatory, Williams Bay, Wis.

**Sturtevant, Alfred Henry**, 1930 (8), California Institute of Technology, Pasadena 4, Calif.

**Suits,** Chauncey Guy, 1946 (4), Research Laboratory, General Electric Company, Schenectady 5, N. Y.

**Sumner, James Batcheller**, 1948 (5), New York State College of Agriculture, Laboratory of Enzyme Chemistry, Cornell University, Ithaca, N. Y.

**Sverdrup, Harald Ulrik**, 1945 (3), Norsk Polarinstitutt, Observatoriegt. 1, Oslo, Norway

**Swanton, John** Reed, 1932 (11), 22 George Street, Newton 58, Mass.

**Taliaferro, William** Hay, 1940 (8), University of Chicago, Chicago 37, Ill.

**Tate, John** Torrence, 1942 (3), University of Minnesota, Minneapolis 14, Minn.

**Teller, Edward,** 1948 (3), Institute for Nuclear Studies, University of Chicago, Chicago 37, Ill.

**Terman,** Frederick Emmons, 1946 (4), School of Engineering, Stanford University, Stanford University, Calif.

**Terman, Lewis** Madison, 1928 (12), 761 Dolores Street, Stanford University, Calif.

**Thimann, Kenneth** Vivian, 1948 (7), Harvard University, Biological Laboratories, 16 Divinity Avenue, Cambridge 38, Mass.

**Thom, Charles,** 1937 (7), 207 Grant Street, Port Jefferson, N. Y.

**Thomas, Charles Allen,** 1948 (4), Monsanto Chemical Company, 1700 South Second Street, St. Louis 4, Mo.

**Thomas,** Tracy Yerkes, 1941 (1), Swain Hall, Indiana University, Bloomington, Ind.

**Thorndike, Edward** Lee, 1917 (12), Teachers College, Columbia University, New York 27, N. Y.

**Thurstone,** Louis Leon, 1938 (12), University of Chicago, Chicago 37, Ill.

**Timoshenko, Stephen** Prókop, 1940 (4), Room 262, Engineering Building, Stanford University, Calif.

**Tolman, Edward** Chace, 1937 (12), Department of Psychology, University of California, Berkeley 4, Calif.

**Tozzer, Alfred** Marston, 1942 (11), Peabody Museum, Harvard University, Cambridge 38, Mass.

**Trumpler, Robert Julius,** 1932 (2), Berkeley Astronomical Department, University of California, Berkeley 4, Calif.

**Tuve,** Merle Antony, 1946 (3), Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road, N. W., Washington 15, D. C.

**Tyzzer,** Ernest Edward, 1942 (10), 175 Water Street, Wakefield, Mass.

**Urey, Harold** Clayton, 1935 (5), Institute for Nuclear Studies, University of Chicago, Chicago 37, Ill.

**Vandiver,** Harry Shultz, 1934 (1), Department of Applied Mathematics, University of Texas, Austin 12, Tex.

**Van Niel,** Cornelis Bernardus, 1945 (7), Hopkins Marine Station of Stanford University, Pacific Grove, Calif.

**Van Slyke, Donald** Dexter, 1921 (9), Rockefeller Institute for Medical Research, 66th Street and York Avenue, New York 21, N. Y.

**Van Vleck,** John **Hasbrouck,** 1935 (3), Harvard University, Cambridge 38, Mass.

**Vaughan,** Thomas **Wayland,** 1921 (6), 3333 P Street, N. W., Washington 7, D. C.

**Veblen, Oswald,** 1919 (1), The Institute for Advanced Study, Princeton, N. J.

**Vickery,** Hubert Bradford, 1943 (9), Connecticut Agricultural Experiment Station, New Haven 4, Conn.

**von Karman,** Theodore, 1938 (4), California Institute of Technology, Pasadena 4, Calif.

**von Neumann, John,** 1937 (1), The Institute for Advanced Study, Princeton, N. J.

**Waksman, Selman** Abraham, 1942 (7), Agricultural Experiment Station, New Brunswick, N. J.

**Walker,** John Charles, 1945 (7), 206 Horticulture Building, University of Wisconsin, Madison 6, Wis.

**Walsh, Joseph** Leonard, 1936 (1), Harvard University, Cambridge 38, Mass.

**Webster,** David Locke, 1923 (3), Physics Department, Room 385, Stanford University, Stanford University, Calif.

**Weiss, Paul** Alfred, 1947 (8), Department of Zoology, University of Chicago, Chicago 37, Ill.

**Went,** Frits Warmolt, 1947 (7), California Institute of Technology, Pasadena 4, Calif.

**Werkman,** Chester Hamlin, 1946 (9), Department of Bacteriology, Iowa State College, Ames, Iowa

**Wetmore, Alexander,** 1945 (8), Smithsonian Institution, Washington 25, D. C.

**Wever,** Ernest Glen, 1940 (12), Princeton University, Princeton, N. J.

**Weyl,** Claus Hugo **Hermann,** 1940 (1), The Institute for Advanced Study, Princeton, N. J.

**Whipple, George** Hoyt, 1929 (10), School of Medicine and Dentistry, University of Rochester, 260 Crittenden Boulevard, Rochester 7, N. Y.

**Whitehead,** John Boswell, 1932 (4), Johns Hopkins University, Baltimore 18, Md.

**Whitney, Hassler,** 1945 (1), Eliot House N21, Cambridge 38, Mass.

**Whitney,** Willis Rodney, 1917 (5), General Electric Company, Schenectady 5, N. Y.

**Wigner, Eugene Paul,** 1945 (3), 8 Ober Road, Princeton, N. J.

**Williams, Robert R.,** 1945 (5), 297 Summit Avenue, Summit, N. J.

**Williams, Roger** John, 1946 (5), Biochemical Institute, University of Texas, Austin 12, Tex.

**Willier, Benjamin** Harrison, 1945 (8), Department of Biology, Johns Hopkins University, Baltimore 18, Md.

**Wilson, Edgar Bright, Jr.,** 1947 (5), Harvard University, Department of Chemistry, 12 Oxford Street, Cambridge 38, Mass.

**Wilson, Edwin** Bidwell, 1919 (3), Harvard School of Public Health, 695 Huntington Avenue, Boston 15, Mass.

**Wilson, Robert** Erastus, 1947 (4), 910 South Michigan Avenue, Chicago 80, Ill.

**Wislocki, George** Bernays, 1941 (8), Harvard Medical School, 25 Shattuck Street, Boston 15, Mass.

**Wolbach,** Simeon **Burt,** 1938 (10), Department of Pathology, Children's Hospital, 300 Longwood Avenue, Boston 15, Mass.

**Wood,** Robert Williams, 1912 (3), Johns Hopkins University, Baltimore 18, Md.

**Woodring,** Wendell **Phillips,** 1946 (6), United States Geological Survey, Washington 25, D. C.

**Woodworth,** Robert Sessions, 1921 (12), Columbia University, New York 27, N. Y.

**Wright,** Frederick Eugene, 1923 (6), 2134 Wyoming Avenue, N. W., Washington 8, D. C.

**Wright, Sewall** Green, 1934 (8), Department of Zoology, University of Chicago, Chicago 37, Ill.

**Wright,** William Hammond, 1922 (2), 60 North Keeble Avenue, San Jose 11, Calif.

**Wulf, Oliver** Reynolds, 1949 (5), Weather Bureau Regional Office, Gates and Crellin Laboratories, California Institute of Technology, Pasadena 4, Calif.

**Wyckoff, Ralph** Walter Graystone, 1949 (5), Laboratory of Physical Biology, National Institute of Health, Bethesda 14, Md.

**Yerkes, Robert** Mearns, 1923 (12), Yale University School of Medicine, 333 Cedar Street, New Haven, Conn.

**Yost, Don** Merlin Lee, 1944 (5), California Institute of Technology, Pasadena 4, Calif.

**Zachariasen,** Frederik **William** Houlder, 1949 (3), Department of Physics University of Chicago, Chicago 37, Ill.

**Zariski, Oscar,** 1944 (1), Department of Mathematics, Harvard University, Cambridge 38, Mass.

**Zworykin,** Vladimir Kosma, 1943 (4), 103 Battle Road, Princeton, N. J.

Number of Members July 1, 1949: 448

## MEMBERS EMERITI

**Benedict, Francis** Gano, 1914, Machiasport, Me.

**Dewey, John,** 1910, 1158 Fifth Avenue, New York 29, N. Y.

**Dickson,** Leonard Eugene, 1913, Route #2, Joliet, Ill.

**Fernald,** Merritt Lyndon, 1935, Gray Herbarium, Harvard University, Cambridge 38, Mass.

**Hektoen, Ludvig,** 1918, Chicago Tumor Institute, 21 West Elm Street, Chicago 10, Ill.

**Howard,** Leland **Ossian,** 1916, Bureau of Entomology, U. S. Department of Agriculture, Washington 25, D. C.

**Stratton,** George **Malcolm,** 1928, University of California, Berkeley 4, Calif.

## FOREIGN ASSOCIATES

The letter in parentheses following the year of election indicates the field of scientific research in which the foreign associate was working at the time of his election, as follows:

(A) Mathematics
(B) Astronomy
(C) Physics
(D) Engineering
(E) Chemistry
(F) Geology and Paleontology
(G) Botany
(H) Zoology and Anatomy
(I) Physiology and Biochemistry
(J) Pathology and Bacteriology
(K) Anthropology and Psychology

**Adrian, Edgar Douglas,** 1941 (I), Trinity College, Cambridge, England
**Alexandroff, Paul A.,** 1947 (A), Mathematical Institute of the Academy of Sciences of the U.S.S.R., Bolshaya Kalushskaya 19, Moscow, U.S.S.R.
**Bailey, Sir Edward,** 1944 (F), 19 Greenhill Gardens, Edinburgh 10, Scotland
**Bartlett, Sir Frederic** Charles, 1947 (K), University of Cambridge, The Psychological Laboratory, Downing Place, Cambridge, England
**Bjerknes, V. (F. K.),** 1934 (F), The University, Oslo, Norway
**Bohr, Niels,** 1925 (C), Institute for Theoretical Physics, Blegdamsvej 15, Copenhagen, Denmark
**Bordet, Jules,** 1935 (I), Pasteur Institute, Rue du Remorqueur, 28, Brussels, Belgium
**Bragg,** Sir William **Lawrence,** 1945 (C), Cavendish Laboratory, The University, Cambridge, England
**de Broglie, Prince Louis,** 1948 (C), 94 Rue Perronet, Neuilly-sur-Seine, France
**Cartan, Élie,** 1949 (A), University of Paris, Paris, France
**Caso, Alfonso,** 1943 (K), Secretaria de Bienes Nacionales e Inspeccion Administrativa, Mexico, D. F.
**Chapman, Sydney,** 1946 (A), Letters: Queen's College, Oxford; Printed matter: The Mathematical Institute, Oxford, England
**Dale, Sir Henry Hallett,** 1940 (I), The Wellcome Trust, 28 Portman Square, London, W.1, England
**Debye, Peter,** 1931* (C), Baker Laboratory, Cornell University, Ithaca, N. Y. (U.S.A.)
**Dirac, Paul** Adrien **Maurice,** 1949 (A), Department of Mathematics, St. John's College, Cambridge, England
**Einstein, Albert,** 1922† (C), The Institute for Advanced Study, Princeton, N. J. (U.S.A.)

---

[*] Dr. Debye became a naturalized citizen in 1946 and a member of the Academy in 1947.

[†] Dr. Einstein became a naturalized citizen in 1941 and a member of the Academy in 1942.

**Fisher,** Ronald Aylmer, 1948 (H), Department of Genetics, University of Cambridge, 44 Storey's Way, Cambridge, England

**Hadamard, Jacques,** 1926 (A), 12, rue Emile Faguet, Paris XIV, France

**Helland-Hansen, Björn,** 1947 (F), Chr. Michelsens Institutt for Videnscap, Bergen, Norway

**Hill, Archibald Vivian,** 1941 (I), 16 Bishopwood Road, Highgate, London, N.6, England

**Hill, James Peter,** 1940 (H), Kanimbla, Dollis Avenue, London N.3, England

**Houssay, Bernardo Alberto,** 1940 (I), Viamonte 2790, Buenos Aires, Argentina

**Jeffreys, Harold,** 1945 (B), St. John's College, Cambridge, England

**Jones, Sir** Harold **Spencer,** 1943 (B), Royal Observatory, Greenwich, London, S.E.10, England

**Kapitza, Peter** Leonidovich, 1946 (C), Institute for Physical Problems, Academy of Sciences of the U.S.S.R., Moscow, U.S.S.R.

**Karrer, Paul,** 1945 (E), University of Zurich, Zurich, Switzerland

**Keith, Sir Arthur,** 1941 (I, K), Buckston Browne Farm, Downe, Farnborough, Kent, England

**Krogh, August,** 1937 (H), The Laboratory, Söbredden, Gentofte, Denmark

**Levi, Giuseppe,** 1940 (H), Instituto di Anatomia Umana, Corso Massimo D'Azeglio, 52, Turin, Italy

**Lim, Robert K. S.,** 1942 (I), Army Medical Administration, Municipal Government, Shanghai, China

**Linderstrøm-Lang,** Kaj Ulrik, 1947 (E), Chemical Division, Carlsberg Laboratory, Copenhagen, Denmark

**Lyot,** Bernard Ferdinand, 1949 (B), 9 bis rue Boileau, Paris 16, France

**Pieron, Henri,** 1949 (K), Institute of Psychology, University of Paris, Paris, France

**Robinson, Sir Robert,** 1934 (E), Dyson Perrins Laboratory, South Parks Road, Oxford, England

**Ruzicka, Leopold,** 1944 (E), Department of Organic Chemistry, Institute of Technology, Zurich, Switzerland

**Sherrington, Sir Charles Scott,** 1924 (I), Gonville and Caius College, Cambridge, England

**Sommerfeld, Arnold,** 1929 (C), Dunant-strasse 6, Munich 23, Germany

**Southwell, Richard** Vynne, 1943 (D), Imperial College of Science and Technology, South Kensington, London, S.W.7, England

**Svedberg, The,** 1945 (E), Fysikalisk-Kemiska Institutionen, University of Uppsala, Uppsala, Sweden

**Taylor, Sir Geoffrey I.,** 1945 (A), Trinity College, Cambridge, England

**Tiselius, Arne** W. K., 1949 (I), Institute of Biochemistry, Uppsala University, Uppsala, Sweden

**Vallee-Poussin, C. de la,** 1929 (A), University of Louvain, Louvain, Belgium

**Vening Meinesz, Felix Andries,** 1939 (F), Potgieterlaan 5, Amersfoort, Holland

**Watson, D. M. S.,** 1938 (H), University College, Gower Street, London, W.C.1, England

**Wieland, Heinrich,** 1932 (E), Sophienstrasse 9, Munich 2 NW, Germany

**Winge, Öjvind,** 1949 (I), Department of Physiology, Carlsberg Laboratory, Copenhagen (Valby), Denmark

**Yukawa, Hideki,** 1949 (C), The Institute for Advanced Study, Princeton, N. J. After Oct. 1, 1949: Department of Physics, Columbia University, New York 27, N. Y. (U. S. A.)

Number of Foreign Associates July 1, 1949: 45.

If a foreign associate becomes a member of the Academy his name is not counted in the limit of 50 foreign associates.

## SECTIONS

### (1) *Mathematics*—27 members

Zariski, Oscar, *Chairman* (1952)
Albert, A. A.
Alexander, J. W.
Bell, E. T.
Bliss, G. A.
Coble, A. B.
Douglas, Jesse
Eisenhart, L. P.
Evans, G. C.
Kasner, Edward
Lefschetz, Solomon
McShane, E. J.
Mac Lane, Saunders
Miller, G. A.
Moore, R. L.
Morse, Marston
Murnaghan, F. D.
Ritt, J. F.
Smith, Paul A.
Stone, M. H.
Thomas, T. Y.
Vandiver, H. S.
Veblen, Oswald
von Neumann, John
Walsh, J. L.
Weyl, Hermann
Whitney, Hassler

### (2) *Astronomy*—27 members

Mitchell, S. A., *Chairman* (1950)
Abbot, C. G.
Adams, W. S.
Aitken, R. G.
Anderson, J. A.
Babcock, H. D.
Bowen, I. S.
Fleming, J. A.
Hubble, E. P.
Joy, A. H.
King, A. S.
Lambert, W. D.
Leuschner, A. O.
Mayall, N. U.
Menzel, D. H.
Merrill, P. W.
Moulton, F. R.
Nicholson, S. B.
Ross, F. E.
Russell, H. N.
Seares, F. H.
Shapley, Harlow
Slipher, V. M.
Stebbins, Joel
Struve, Otto
Trumpler, R. J.
Wright, W. H.

### (3) *Physics*—65 members

Beams, J. W., *Chairman* (1951)
Allison, S. K.
Alvarez, L. W.
Anderson, C. D.
Bacher, R. F.
Bainbridge, K. T.
Berkner, L. V.
Bethe, H. A.
Birge, R. T.
Bjerknes, Jacob
Bloch, Felix
Breit, Gregory
Bridgman, P. W.
Brode, R. B.
Coblentz, W. W.
Compton, A. H.
Compton, K. T.
Condon, E. U.
Coolidge, W. D.
Crew, Henry
Davis, Bergen
Davisson, C. J.
Dempster, A. J.
DuBridge, L. A.
Dunning, J. R.
Einstein, Albert

Epstein, P. S.
Fermi, Enrico
Franck, James
Goudsmit, S. A.
Houston, W. V.
Hull, A. W.
Ives, H. E.
Kemble, E. C.
Lauritsen, C. C.
Lawrence, E. O.
Loomis, F. W.
Lyman, Theodore
McMillan, E. M.
Mason, Max
Millikan, R. A.
Mulliken, R. S.
Oppenheimer, J. R.
Pegram, G. B.
Pierce, G. W.
Piggot, C. S.
Rabi, I. I.
Reichelderfer, F. W.
Rossby, C.-G.
Saunders, F. A.
Schwinger, Julian
Slater, J. C.
Slichter, L. B.
Stern, Otto
Stewart, G. W.
Sverdrup, H. U.
Tate, J. T.
Teller, Edward
Tuve, M. A.
Van Vleck, J. H.
Webster, D. L.
Wigner, E. P.
Wilson, Edwin B.
Wood, R. W.
Zachariasen, W. H.

### (4) *Engineering*—35 members

Dryden, H. L., *Chairman* (1950)
Adams, C. A.
Briggs, L. J.
Buckley, O. E.
Bush, Vannevar
Cochrane, E. L.
Curme, G. O., Jr.
Dunn, Gano
Durand, W. F.
Fletcher, Harvey
Foote, P. D.
Gibbs, W. F.
Gilliland, E. R.
Herty, C. H., Jr.
Hoover, Herbert
Hovgaard, William
Hunsaker, J. C.
Jeffries, Zay
Jewett, F. B.
Kelly, M. J.
Kettering, C. F.
Lewis, W. K.
Loomis, A. L.
Merica, P. D.
Savage, J. L.
Slepian, Joseph
Soderberg, C. R.
Suits, C. G.
Terman, F. E.
Thomas, C. A.
Timoshenko, Stephen
von Karman, T.
Whitehead, J. B.
Wilson, R. E.
Zworykin, V. K.

### (5) *Chemistry*—62 members

Latimer, W. M., *Chairman* (1950)
Adams, L. H.
Adams, Roger
Adkins, Homer
Bachmann, W. E.
Bancroft, W. D.
Bartlett, P. D.
Baxter, G. P.
Bogert, M. T.
Bolton, E. K.
Conant, J. B.
Cope, A. C.
Daniels, Farrington
Debye, Peter
Elderfield, R. C.
Eyring, Henry
Fieser, L. F.
Folkers, Karl
Fuson, R. C.
Giauque, W. F.
Gilman, Henry
Hammett, L. P.
Harkins, W. D.
Hildebrand, J. H.
Hudson, C. S.
Hulett, G. A.
Ipatieff, V. N.
Jacobs, W. A.
Johnson, J. R.
Keyes, F. G.
Kharasch, M. S.
Kirkwood, J. G.
Kistiakowsky, G. B.
Kraus, C. A.
La Mer, V. K.
Lamb, A. B.
Langmuir, Irving
Lind, S. C.
Longsworth, L. G.
McElvain, S. M.
MacInnes, D. A.
Marvel, C. S.
Mayer, J. E.
Noyes, W. A., Jr.
Onsager, Lars
Pauling, Linus
Pitzer, K. S.
Rodebush, W. H.
Scatchard, George
Schlesinger, H. I.

Seaborg, G. T.
Small, L. F.
Smith, L. I.
Sumner, J. B.
Urey, H. C.
Whitney, W. R.
Williams, Robert R.
Williams, Roger J.
Wilson, E. Bright, Jr.
Wulf, O. R.
Wyckoff, R. W.
Yost, D. M.

### (6) *Geology*—37 members

Rubey, W. W., *Chairman* (1951)
Allen, E. T.
Berkey, C. P.
Blackwelder, Eliot
Bowen, N. L.
Bowman, Isaiah
Bradley, W. H.
Bucher, W. H.
Buddington, A. F.
Byerly, Perry
Chaney, R. W.
Daly, R. A.
Day, A. L.
Dunbar, C. O.
Ewing, Maurice
Gilluly, James
Gregory, W. K.
Gutenberg, Beno
Hewett, D. F.
Kelley, W. P.
Knopf, Adolph
Larsen, E. S., Jr.
Lawson, A. C.
Leith, C. K.
Longwell, C. R.
Lovering, T. S.
Macelwane, J. B.
Mead, W. J.
Mendenhall, W. C.
Palache, Charles
Reeside, J. B., Jr.
Ruedemann, Rudolf
Simpson, G. G.
Stock, Chester
Vaughan, T. W.
Woodring, W. P.
Wright, F. E.

### (7) *Botany*—37 members

Cleland, R. E., *Chairman* (1950)
Allen, C. E.
Babcock, E. B.
Bailey, I. W.
Bailey, L. H.
Beadle, G. W.
Blakeslee, A. F.
Brink, R. A.
Burkholder, P. R.
Campbell, D. H.
Chandler, W. H.
Couch, J. N.
Delbrück, Max
Dodge, B. O.
Duggar, B. M.
Fred, E. B.
Hoagland, D. R.
Jones, D. F.
Kunkel, L. O.
McClintock, Barbara
Mangelsdorf, P. C.
Merrill, E. D.
Osterhout, W. J. V.
Raper, K. B.
Rhoades, M. M.
Robbins, W. J.
Sax, Karl
Sinnott, E. W.
Smith, Gilbert M.
Stadler, L. J.
Stakman, E. C.
Thimann, K. V.
Thom, Charles
Van Niel, C. B.
Waksman, S. A.
Walker, J. C.
Went, F. W.

### (8) *Zoology and Anatomy*—39 members

Romer, Alfred S., *Chairman* (1952)
Bartelmez, G. W.
Bigelow, H. B.
Castle, W. E.
Child, C. M.
Conklin, E. G.
Corner, G. W.
Danforth, C. H.
Demerec, Milislav
Detwiler, S. R.
Dobzhansky, Theodosius
Dunn, L. C.
Goldschmidt, R. B.
Harrison, R. G.
Hartman, C. G.
Harvey, E. N.
Herrick, C. J.
Hisaw, F. L.
Jacobs, M. H.
Lewis, W. H.
Metz, C. W.
Moore, C. R.
Muller, H. J.
Nicholas, J. S.
Painter, T. S.
Parker, G. H.
Patterson, J. T.
Riddle, Oscar
Schmitt, F. O.
Smith, Philip E.
Sonneborn, T. M.

Stern, Curt
Sturtevant, A. H.
Taliaferro, W. H.
Weiss, Paul
Wetmore, Alexander
Willier, B. H.
Wislocki, G. B.
Wright, Sewall

### (9) *Physiology and Biochemistry*—48 members

Doisy, E. A., *Chairman* (1951)
Anderson, R. J.
Ball, E. G.
Bard, Philip
Bronk, D. W.
Carlson, A. J.
Clark, W. M.
Clarke, H. T.
Cohn, E. J.
Cori, Carl F.
Cori, Gerty T.
Davis, Hallowell
DuBois, E. F.
du Vigneaud, Vincent
Elvehjem, C. A.
Erlanger, Joseph
Evans, H. M.
Fenn, W. O.
Forbes, Alexander
Gasser, H. S.
Hart, E. B.
Hartline, H. K.
Hastings, A. B.
Lewis, H. B.
Link, K. P.
Loeb, R. F.
Long, C. N. H.
McCollum, E. V.
MacNider, W. deB.
Marshall, E. K., Jr.
Maynard, L. A.
Meek, W. J.
Meyerhof, Otto
Michaelis, Leonor
Northrop, J. H.
Oncley, J. L.
Peters, J. P.
Richards, A. N.
Rose, W. C.
Schmidt, C. F.
Shaffer, P. A.
Sherman, H. C.
Smith, Homer W.
Stadie, W. C.
Stanley, W. M.
Van Slyke, D. D.
Vickery, H. B.
Werkman, C. H.

### (10) *Pathology and Bacteriology*—38 members

Dochez, A. R., *Chairman* (1951)
Addis, Thomas
Armstrong, Charles
Avery, O. T.
Blake, F. G.
Blalock, Alfred
Cannon, P. R.
Castle, W. B.
Coggeshall, L. T.
Cole, Rufus
Dubos, R. J.
Francis, Thomas, Jr.
Gamble, J. L.
Goodpasture, E. W.
Graham, E. A.
Heidelberger, Michael
Horsfall, F. L., Jr.
Huggins, C. B.
Kelser, R. A.
Little, C. C.
Loeb, Leo
Long, E. R.
Longcope, W. T.
Meyer, K. F.
Minot, G. R.
Mueller, J. Howard
Murphy, J. B.
Novy, F. G.
Opie, E. L.
Paul, J. R.
Rivers, T. M.
Robertson, O. H.
Rous, Peyton
Sabin, Florence R.
Shope, R. E.
Tyzzer, E. E.
Whipple, G. H.
Wolbach, S. B.

### (11) *Anthropology*—10 members

Linton, Ralph, *Chairman* (1951)
Hooton, E. A.
Kidder, A. V.
Kroeber, A. L.
Lowie, R. H.
Schultz, A. H.
Shapiro, H. L.
Spier, Leslie
Swanton, J. R.
Tozzer, A. M.

### (12) *Psychology*—23 members

Boring, E. G., *Chairman* (1950)
Beach, F. A.
Carmichael, Leonard
Gesell, Arnold
Graham, C. H.
Hilgard, E. R.
Hull, C. L.
Hunter, W. S.
Köhler, Wolfgang
Lashley, K. S.
Miles, W. R.
Pillsbury, W. B.
Richter, C. P.
Seashore, C. E.
Stevens, S. S.
Stone, C. P.
Terman, L. M.
Thorndike, E. L.
Thurstone, L. L.
Tolman, E. C.
Wever, E. G.
Woodworth, R. S.
Yerkes, R. M.

### *Temporary Nominating Group on Geophysics*—1942–51

30 members

John A. Fleming, *Chairman*

*Mathematics:* F. D. Murnaghan.

*Astronomy:* C. G. Abbot, J. A. Fleming, Walter D. Lambert, Harlow Shapley.

*Physics:* L. V. Berkner, Jacob Bjerknes, P. W. Bridgman, A. H. Compton, C. S. Piggot, F. W. Reichelderfer, C.-G. Rossby, L. B. Slichter, H. U. Sverdrup.

*Engineering:* L. J. Briggs, J. C. Hunsaker.

*Chemistry:* L. H. Adams, S. C. Lind, Oliver R. Wulf.

*Geology:* Isaiah Bowman, Perry Byerly, Arthur L. Day, Maurice Ewing, Beno Gutenberg, Adolph Knopf, J. B. Macelwane, T. Wayland Vaughan, F. E. Wright.

*Botany:* S. A. Waksman.

*Zoology and Anatomy:* H. B. Bigelow.

### *Temporary Nominating Group on Medicine*—1949

23 members

Warfield T. Longcope, *Chairman*

*Zoology and Anatomy:* George W. Corner, William H. Taliaferro, George B. Wislocki.

*Physiology and Biochemistry:* A. J. Carlson, Eugene F. DuBois, Wallace O. Fenn, Herbert S. Gasser, C. N. H. Long, W. deB. MacNider, Walter J. Meek, Homer W. Smith, William C. Stadie.

*Pathology and Bacteriology:* Francis G. Blake, W. B. Castle, L. T. Coggeshall, Evarts A. Graham, Michael Heidelberger, Charles Huggins, Warfield T. Longcope, Thomas M. Rivers, O. H. Robertson, Peyton Rous, George H. Whipple.

## COMMITTEES

### *Auditing*

John B. Reeside, Jr., *Chairman;* Charles S. Piggot, Merle A. Tuve.

### *Biographical Memoirs*

Alfred N. Richards, *Chairman, ex officio,* President of the Academy.
Chairmen of Sections of the Academy.

### *Buildings and Grounds Advisory Committee*

(Joint Committee of the Academy and Research Council)

G. D. Meid, *Chairman;* Detlev W. Bronk (NAS), R. C. Gibbs (NRC), L. H. Weed (NRC), F. E. Wright (NAS).

### *Exhibits*

(Joint Committee of the Academy and Research Council)

F. E. Wright, *Chairman;* Harlow Shapley, member-at-large.
Chairmen of Sections of the Academy.
Chairmen of Divisions of the National Research Council.
Raymund L. Zwemer, *Secretary.*

### *Finance*

W. J. Robbins, *Chairman, ex officio,* Treasurer of the Academy.
Alfred N. Richards, *ex officio,* President of the Academy.
Detlev W. Bronk, *ex officio,* Chairman of the National Research Council.
Vannevar Bush, L. P. Eisenhart, J. C. Hunsaker.

### *Financial Advisory*

Lindsay Bradford, Walter S. Gifford, Henry Morgan.
Horace Ford, *Financial Adviser.*

### *Library*

G. W. Corner, *Chairman;* Lloyd V. Berkner, Hugh L. Dryden, James L. Gamble.

### *Standing Committee on Meetings*

F. E. Wright, *Chairman* (1950); Edwin G. Boring (1951), Ralph E. Cleland (1951), John A. Fleming (1950), Harvey Fletcher (1952), Ernest W. Goodpasture (1952), Joel H. Hildebrand (1951), Duncan A. MacInnes (1952), Edwin B. Wilson (1950).

### *Advisory Committee on Membership*

Members of the Council of the Academy, Chairmen of Sections of the Academy.

### *Publications of the Academy*

Alfred N. Richards, *ex officio,* President of the Academy; F. E. Wright, *ex officio,* Home Secretary of the Academy; E. G. Conklin.

### *Retirement Pension Trust Agreement*

*Pension Committee:* Alfred N. Richards, President of the Academy
William J. Robbins, Treasurer of the Academy
F. E. Wright, Home Secretary of the Academy
Detlev W. Bronk, Chairman of National Research Council

*Trustees:* G. D. Meid, Alfred N. Richards.

### *Revision of the Constitution*

Edwin B. Wilson, *Chairman;* Eugene F. DuBois, William J. Robbins.

### *Weights, Measures, and Coinage*

Gano Dunn, *Chairman;* G. P. Baxter, R. W. Wood.

## TRUST FUNDS

### *Alexander Dallas Bache Fund*

Researches in physical and natural science. $60,000.

*Board of Directors:* Edwin B. Wilson, *Chairman;* W. J. V. Osterhout, Sewall Wright.

### *Barnard Medal for Meritorious Service to Science*

Discoveries in physical, or astronomical science or novel application of science to purposes beneficial to the human race. (Every five years the committee recommends the person whom they consider the most deserving of the medal, and upon approval by the Academy, the name of the nominee is forwarded to the trustees of Columbia University, who administer the Barnard Medal Fund.)

*Committee:* Edwin Hubble, *Chairman;* Bergen Davis, Ernest O. Lawrence, V. M. Slipher, W. R. Whitney.

### *John J. Carty Fund*

Medal and monetary award, not oftener than once in every two years, to an individual for noteworthy and distinguished accomplishment in any field of science coming within the scope of the charter of the Academy. $25,000.

*Committee:* Herbert Gasser, *Chairman* (1952); Roger Adams (1954), Vannevar Bush (1950), Harvey Fletcher (1951), Robert E. Wilson (1953).

### *Cyrus B. Comstock Fund*

Prize awarded every five years for most important discovery or investigation in electricity, magnetism, and radiant energy; or to aid worthy investigation in those subjects. $15,000.

*Committee:* I. I. Rabi, *Chairman* (1952); J. W. Beams (1951), Arthur H. Compton (1954), Lee A. DuBridge (1950), Enrico Fermi (1953).

### *Henry Draper Fund*

Medal and investigations in astronomical physics. $10,000.

*Committee:* I. S. Bowen, *Chairman* (1952); W. S. Adams (1950), J. A. Fleming (1951), P. W. Merrill (1953), Henry Norris Russell (1954).

### *Daniel Giraud Elliot Fund*

Medal and honorarium for most meritorious work in zoology or paleontology published each year. $8,000.

*Committee:* Alexander Wetmore, *Chairman* (as Secretary, Smithsonian Institution); Albert E. Parr (as Scientific Director, American Museum of Natural History); Alfred S. Romer (1950), A. H. Sturtevant.

### *Wolcott Gibbs Fund*

Chemical research. $10,545.50.

*Board of Directors:* G. P. Baxter, *Chairman;* Roger Adams, C. A. Kraus.

### *Benjamin Apthorp Gould Fund*

Researches in astronomy. $30,000.

*Board of Directors:* F. R. Moulton, *Chairman;* W. S. Adams, Joel Stebbins.

### *Marcellus Hartley Fund*

Medal for eminence in the application of science to the public welfare. $1,200.

*Committee:* W. V. Houston, *Chairman* (1950); Isaiah Bowman (1951), Oliver E. Buckley (1952), Lee A. DuBridge (1951), H. J. Muller (1950), J. R. Oppenheimer (1952).

### *Joseph Henry Fund*

To assist meritorious investigators, especially in the direction of original research. $47,163.50.

*Committee:* W. S. Hunter, *Chairman* (1950); G. W. Corner (1952), E. B. Fred (1953), A. Baird Hastings (1954), Joel H. Hildebrand (1951).

### *Marsh Fund*

Original research in the natural sciences. $20,000.

*Committee:* Edmund W. Sinnott, *Chairman* (1950); A. F. Buddington (1951), C. S. Marvel (1954), W. M. Stanley (1953), W. H. Taliaferro (1952).

### *Murray Fund*

Agassiz Medal for original contribution in the science of oceanography. $10,000.

*Committee:* C. S. Piggot, *Chairman* (1950); Maurice Ewing (1952), W. W. Rubey (1951).

### *J. Lawrence Smith Fund*

Medal and investigations of meteoric bodies. $10,000.

*Committee:* John A. Fleming, *Chairman* (1952), Harlow Shapley (1953) V. M. Slipher (1950), Joel Stebbins (1954), R. J. Trumpler (1951).

### *Mary Clark Thompson Fund*

Medal for most important services to geology and paleontology. $10,000.

*Committee:* A. F. Buddington, *Chairman* (1950); Ralph W. Chaney (1952), G. G. Simpson (1951).

### *Charles Doolittle Walcott Fund*

Medal and honorarium to stimulate research in pre-Cambrian or Cambrian life. $5,000.

*Board of Trustees:* Carl O. Dunbar, *Chairman;* Alexander Wetmore, *ex officio*, Secretary, Smithsonian Institution; Pierre Pruvost, representing the Institut de France (1952); C. J. Stubblefield, representing The Royal Society of London (1952); John B. Reeside, Jr.

### *James Craig Watson Fund*

Medal and the promotion of astronomical research. $25,000.

*Trustees:* A. O. Leuschner, *Chairman;* F. E. Ross, R. J. Trumpler.

PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES

Volume 35 August 15, 1949 Number 8

---

# A METHOD OF EFFECTING ZYGOSPORE GERMINATION IN CERTAIN CHLOROPHYCEAE

BY RICHARD C. STARR

DEPARTMENT OF BIOLOGY, VANDERBILT UNIVERSITY

Communicated by E. W. Sinnott, June 11, 1949

*Introduction.*—The potential advantages of genetical studies of haplobiontic Chlorophyceae having zygotic meiosis have in many cases not been realized due to the difficulty of obtaining consistently high percentages of zygospore germination. There are comparatively few reports in the literature in which successful methods of germination are reported. Bold[1] summarized the efforts of those investigators who achieved some success by simulating in culture the environmental conditions which algal zygospores normally encounter in nature. Among these were included the transfer of zygospores to fresh media, alternate drying and wetting and other methods. Rowan,[2] using arsenic, strychnine and indole acetic acid, had considerable success with the germination of the zygospores of *Hydrodictyon*. Pocock[3] was able to germinate the zygospores of *Volvox* five or six months after their formation by placing them in dilute culture solution in direct sunlight. She concluded that direct sunlight was essential for germination, and that complete drying of the zygospores before introducing them into the culture medium was beneficial though not necessary. Lewin,[4] working with a strain of *Chlamydomonas*, has achieved up to 30% germination of the zygospores by using a method of metabolic starvation. He also concluded that complete darkness and not sunlight was essential for successful germination of the zygospores of his organism. Moewus[5] secured excellent results with *Chlamydomonas eugametos* zygospores by using a soil solution prepared without heat. Metzner[6] was able to secure only 10% germination of the oöspores of *Volvox Carteri* using Moewus' cold water extract, but was able to secure 90% germination using the synthetic medium of Uspenski and Uspenskaja.[7] Metzner was able to germinate only moist spores, although Pocock had reported previously that drying was beneficial in germinating the oöspores of another species of *Volvox*.

*Experimental.*—Inasmuch as the writer has obtained consistently high percentages of zygospore germination in a homothallic organism,[8] similar to *Chlorococcum*, and also with *Chlamydomonas chlamydogama* Bold, by a simple technique, the following notes as to the method used are reported herewith in the hope that they may prove effective for the zygospores of other Chlorophyceae.

S11C-6 is a homothallic *Chlorococcum*-like alga producing gametes which fuse in pairs to form small spherical zygotes usually about 7.5 microns in diameter. After formation these zygotes grow for a period of time varying from 5 to 14 days during which they may increase up to 25 microns in diameter. After this growth period they secrete a thick, reticulate, outer wall. As the zygospores age, they seem to become filled with large oil droplets; however, they retain their green color.

TABLE 1

ZYGOSPORE GERMINATION IN RELATION TO INCUBATION AT VARIOUS TEMPERATURES

| LENGTH OF EXPOSURE IN HOURS | PERCENTAGE OF GERMINATION AT INCUBATION TEMPERATURES OF: 37°C. | 45°C. | 55°C. |
|---|---|---|---|
| 72 | 70% | .. | .. |
| 60 | 90% | .. | .. |
| 48 | 90% | .. | .. |
| 36 | 30% | .. | .. |
| 24 | 30% | .. | .. |
| 20 | 20% | .. | .. |
| 16 | 20% | .. | .. |
| 12 | 0 | .. | .. |
| 8 | 0 | .. | .. |
| 4 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 |
| 1 | 0 | 30% | 30% |
| 1/2 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 |

The alga was grown in 50-ml. Erlenmeyer flasks containing 30 ml. of a modified sterile Bristol's solution[9] to which 0.03 ml. of Armour's crude liver liquid extract had been added. Although the alga will grow in the purely inorganic Bristol's solution, the growth rate is definitely increased with the addition of the liver extract. The cultures were illuminated by a fluorescent light of approximately 50 foot-candle intensity.

In preliminary attempts to obtain germination, zygospores from a month-old culture were transferred to eight small Petri dishes containing Bristol's agar with 0.15% Difco yeast extract. No germination had occurred 16 days later at which time one of the dishes was transferred to a 37°C. incubator for 48 hours. Forty-eight hours after the removal of the Petri dish from the incubator approximately 70% germination was observed in the dish which had been incubated, but no evidence of germination was observed in the unincubated dishes. The above procedure was repeated with another of the remaining dishes and the same results were obtained. No germination of the zygospores was observed in the unincubated dishes until 35 days after their transfer to the fresh medium when an occasional one could be found germinating.

In order to verify these results, a series of Petri dish cultures was prepared using the zygospores from the same culture of S11C-6, now however, approximately 60 days old. Zygospores were again transferred to Bristol's yeast agar in small Petri dishes and incubated for various periods. The results are summarized in table 1. Percentages are approximate and represent estimates based on microscopic examination.

Maximum germination of the 60-day old zygospores occurred within 48–72 hours after exposure of the zygospores to a temperature of 37°C. for 48–60 hours.

Results comparable to the best previously achieved were also obtained using zygospores from a 15-day old culture. Zygospores from younger cultures are difficult to secure in quantity, inasmuch as they do not appear in abundance with their characteristic thick walls until several weeks after the culture has been started, and until these walls appear, it is difficult to distinguish the zygospores from vegetative cells.

Experiments have also been conducted to determine if heat had a similar effect on the zygospores of *Chlamydomonas chalmydogama* Bold. The cultures were grown in a variety of liquid media including Knop's and Volvox solutions. Zygospores were obtained by mixing the heterothallic strains of the alga in hanging drops. Zygospores of various ages were transferred to Petri dishes containing Bristol's agar and incubated at 37° C. for 48 hours. Approximately 90% germination was observed in all cultures within 48–96 hours after the heat treatment.

*Discussion.*—No attempt is made here to speculate as to the relation of the heat treatment to the physiology of the zygospores or as to its rôle in germination. This stimulatory effect of high temperature suggests the work of Shear and Dodge,[10] who have shown that the ascospores of *Neurospora tetrasperma* can be induced to germinate by a short period of heating. Goddard,[11] working with *Neurospora,* found evidence pointing to the presence of three phases in the rate of respiration of ascospores, corresponding to the stages of dormancy, activation and germination. Goddard and Smith,[12] investigating further the respiration of the ascospores of *Neurospora,* obtained results showing that the marked increase in the respiratory rate and then subsequent germination of the heat-treated ascospores was probably due to the activation of carboxylase by the heat.

Notwithstanding the fact that these investigations of the relation between high temperature and zygospore germination are perhaps not extensive enough to make many conclusive statements as to the exact conditions under which the zygospores of S11C-6 and *Chlamydomonas chlamydogama* Bold will germinate, the writer is of the opinion that the method may form a useful point of departure for other investigators who may be confronted with the problem of zygospore germination. It is hoped that this method of inducing germination may prove especially useful to those concerned with genetical problems in which it would be decidedly advantageous to effect zygospore germination in the shortest possible time and with the greatest possible frequency.

[1] Bold, H. C., *Bot. Rev.*, **8**, 69 (1942).

[2] Rowan, M., Unpublished Thesis, Columbia Univ. (1937).

[3] Pocock, M. A., *Ann. So. Afr. Mus.*, **16**, 523 (1933).

[4] Lewin, R. A., Unpublished manuscript.

[5] Moewus, F., *Zeit. Ind. Abst. Ver.*, **78**, 418 (1940).

[6] Metzner, J., *Bull. Torrey Club*, **72**, 86 (1945).

[7] Uspenski, E. E., and Uspenskaja, W. J., *Zeits. Bot.*, **17**, 273 (1925).

[8] This organism will be referred to in this report as "S11C-6," inasmuch as no final conclusion as to its identity has yet been reached. It was isolated from the soil of San Marcellino, Luzon, Philippine Islands, by Dr. Harold C. Bold of Vanderbilt University.

[9] Bold, H. C., *Bull. Torrey Club*, **76**, 101 (1949).

[10] Shear, C. L., and Dodge, B. O., *J. Agr. Res.*, **54**, 1019 (1927).

[11] Goddard, D. R., *J. Gen. Physiol.*, **19**, 45 (1935).

[12] Goddard, D. R., and Smith, P. E., *Plant Physiol.*, **13**, 241 (1938).

---

# *STUDIES ON THE BIOCHEMICAL GENETICS OF YEAST**

By SEYMOUR POMPER† AND P. R. BURKHOLDER

OSBORN BOTANICAL LABORATORY, YALE UNIVERSITY

Communicated June 21, 1949

The mechanism of inheritance in *Saccharomyces cerevisiae* Hansen has been a subject of controversy, for it has not been clear whether, in this yeast, the characteristic mode of inheritance is Mendelian or non-Mendelian.

The work of Lindegren and Lindegren,[1] and Spiegelman,[2] on the inheritance of adaptive enzyme systems, has pointed to a mechanism of inheritance mediated by cytoplasmic, self-duplicating units. Lindegren and Lindegren[3] have recently presented evidence which was interpreted as a gene-to-gene transfer of components, and for which the concept of a "depletion mutation" was proposed. In the face of hypotheses of such an unusual nature, it becomes worth while to examine the possibility of devising critical experiments to determine what the basic mechanism of inheritance is in this organism.

It is of prime importance in such studies that characters controlled by single genes be studied as criteria by which to judge the genetic process. The technique first used by Beadle and Tatum[4] in *Neurospora,* viz., the induction of biochemically deficient mutants by irradiation, seemed well-suited to this purpose. That such mutants can be produced in yeast has been shown by Reaume and Tatum,[5] who recovered mutants after treatment with nitrogen mustard. It has been amply demonstrated in other organisms, such as *Neurospora,*[6] *Ophiostoma,*[7] and *Escherichia coli,*[8] that the mutant characters, i.e., the biochemical deficiencies, are inherited as though controlled by single genes. Accordingly, haploid strains of *S. cerevisiae* (very generously supplied by Dr. C. C. Lindegren) were treated with ultra-violet irradiation, and mutants were recovered by an appropriately modified screening[9] and detection[10] technique. The mutants that were obtained are listed in table 1.[11] They may be characterized, in general, as being unable to grow in the absence of the required metabolites for at least 72 hours at 30°C.

TABLE 1

BIOCHEMICALLY DEFICIENT MUTANTS OF *S. cerevisiae* PRODUCED BY ULTRA-VIOLET IRRADIATION OF HAPLOID STRAINS

| CULTURE IRRADIATED | NUMBER AND KINDS OF MUTANTS |
|---|---|
| *a** | 1 yeast nucleic acid |
| | 2 methionine |
| | 1 tryptophan |
| | 1 adenine |
| | 1 methionine, adenine, histidine |
| $\alpha$* | 1 unknown factor |
| Tryptophanless *a*† | 1 adenine |
| | 1 uracil |
| | 3 methionine |
| | 3 pantothenicless |
| Methionineless $\alpha$‡ | 3 adenine |
| | 1 uracil |

* Cultures kindly supplied by Dr. C. C. Lindegren.

† Tryptophanless mutant from irradiation of *a*.

‡ Mutant obtained following mustard gas treatment[5] and kindly supplied by Mr. S. E. Reaume, who also gave us a pink, adenineless $\alpha$ haploid.

Lindegren and Lindegren[12] were the first to report a heterothallic condition in *S. cerevisiae.* The symbols "*a*" and "$\alpha$" were given to the major allelic gene pair controlling mating type. These authors[13] devised a mass crossing procedure based on the stability of the heterothallic condition. Winge and Roberts[14] have criticized this procedure on the grounds that aberrant results may be obtained due to selfing. To obviate their objections, we adopted the following procedure: complementarily deficient mutants of opposite mating type were crossed as suggested by Lindegren and Lindegren,[13] and after incubation for 72 hours at 25°C., the cell suspension was plated on agar containing none of the factors that the deficient haploids required. By this means, one recovers only heterozygous, "prototrophic,"[15] diploids, which resemble the wild-type diploid in nutritional requirements. Using mutants deficient for 2 to 3 metabolites, it was possible to rule out statistically the probability of recovering a reverted haploid clone. Since it was possible that a "prototrophic" colony might represent a synergistic association of haploids and diploids, a population analysis was carried out on a prototroph isolate from the cross $a\ Th^-Tr^-Ur_1^-Me_1^+ \times \alpha\ Th^+Tr^+Ur_1^+Me_1^-$.‡ The population, was homogeneously independent as regards nutritional requirements indicating that the single colony isolate represents a pure diploid clone. The final proof was the demonstration of heterozygosity in the segregating ascospores. This over-all mating procedure preserves the advantages of the mass crossing technique, and insures that the material to be examined

TABLE 2

SEGREGATION OF CHARACTERS IN ASCOSPORES FROM THE CROSS $a\ Th^-Ur_1^-Ur_2^+Me_1^+ \times \alpha\ Th^+Ur_1^+Ur_2^-Me_1^-$

| ASCOSPORE | MT | Th | $Me_1$ | Ur | $Ur_1$, $Ur_2$ DISTRIBUTION |
|---|---|---|---|---|---|
| A61.1 | α | − | + | − | $Ur_1^-Ur_2^+$ |
| A61.2 | α | + | + | − | $Ur_1^+Ur_2^-$ |
| A61.3 | *a* | − | − | − | $Ur_1^-Ur_2^-$ |
| A61.4 | *a* | + | − | + | $Ur_1^+Ur_2^+$ |
| A67.1 | *a* | + | − | − | $Ur_1^-Ur_2^+$ |
| A67.2 | α | − | − | − | $Ur_1^+Ur_2^-$ |
| A67.3 | *a* | + | + | − | $Ur_1^-Ur_2^-$ |
| A67.4 | α | − | + | + | $Ur_1^+Ur_2^+$ |
| A68.1 | *a* | − | − | − | $Ur_1^+Ur_2^-$ |
| A68.2 | α | + | − | − | $Ur_1^+Ur_2^-$ |
| A68.3 | α | + | + | − | $Ur_1^-Ur_2^+$ |
| A68.4 | *a* | − | + | − | $Ur_1^-Ur_2^+$ |
| A69.1 | α | − | − | + | $Ur_1^+Ur_2^+$ |
| A69.2 | *a* | + | + | − | $Ur_1^-Ur_2^-$ |
| A69.3 | α | + | − | + | $Ur_1^+Ur_2^+$ |
| A69.4 | *a* | − | + | − | $Ur_1^-Ur_2^-$ |
| A70.1 | α | − | + | − | $Ur_1^+Ur_2^-$ |
| A70.2 | *a* | + | − | − | $Ur_1^+Ur_2^-$ |
| A70.3 | *a* | − | − | − | $Ur_1^-Ur_2^+$ |
| A70.4 | α | + | + | − | $Ur_1^-Ur_2^+$ |

The numbering system used for identification of ascospores is as follows: the letter "A" prefixes the ascus number; a "period" follows the ascus number, and after the "period" comes the ascospore number. Thus A61.1, A61.2, A61.3 and A61.4 refer to the four ascospores of ascus 61. The ascospore numbers are only identification numbers, and have no significance with regard to the sequence of isolation.

genetically has arisen from heterothallic conjugation. This would render unnecessary the precautions of Winge and Laustsen,[16] who carried out crosses of single spores in hanging drop preparations under continual microscopic observation.

Sporulation of prototrophic diploids was induced by the Lindegren and Lindegren[17] modification of the gypsum slant[18] method, and by the procedure of Stantial.[19] The asci analyzed, in all cases here reported, were taken from gypsum slants. The four ascospores of individual asci were separated by microdissection using a modified Fitz micromanipulator.[20] They were then allowed to germinate at 30°C. in hanging drops. The medium (Lindegren, Personal Communication) used was as follows: liquid yeast extract (Anheuser Busch No. 3), 6 ml.; peptone, 5 g.; $MgSO_4 \cdot 7H_2O$, 1 g.; $KH_2PO_4$, 2 g.; sodium lactate (50%), 7.2 g.; dextrose, 40 g.; and water to make 1000 ml. After incubation for 24 hours in hanging drop cultures, the individual haploid clones were transferred, for purposes of further multiplication, to tubes containing nutrient broth, and from these to agar slants, to produce stock cultures.

TABLE 3

RELATION BETWEEN GENOTYPE AND PHENOTYPE IN SEGREGATIONS INVOLVING TWO NON-ALLELIC GENES WITH THE SAME APPARENT PHENOTYPIC EXPRESSION (URACIL REQUIREMENT)

| PHENOTYPIC SEGREGATION (INDEPENDENT:DEPENDENT) | GENOTYPES |
|---|---|
| 0:4 | $Ur_1^-Ur_2^+$, $Ur_1^-Ur_2^+$, $Ur_1^+Ur_2^-$, $Ur_1^+Ur_2^-$ |
| 1:3 | $Ur_1^+Ur_2^+$, $Ur_1^-Ur_2^-$, $Ur_1^+Ur_2^-$, $Ur_1^-Ur_2^+$ |
| 2:2 | $Ur_1^+Ur_2^+$, $Ur_1^+Ur_2^+$, $Ur_1^-Ur_2^-$, $Ur_1^-Ur_2^-$ |

Using this over-all procedure, analyses of several crosses have been carried out. In those cases where only three out of four spores survived, the practice has been followed, for linkage calculations, of assigning the genetic character of the fourth spore by "difference," based on a 2:2 segregation for each single character. This seemed justified in view of the regular results obtained when all four spores survived. In some crosses, the characters of "disperse" and "non-disperse" have been included in the analysis. "Disperse" ($D^+$) refers to homogeneous growth in liquid shake cultures with dextrose as a carbon source in a synthetic medium,[21] and "non-disperse", ($D^-$) signifies flocculence under the same conditions. "Disperse" appears to be dominant over "non-disperse," since the prototrophic diploid arising from the cross "disperse" × "non-disperse" is generally "disperse." However, with sucrose or fructose serving as a carbon source, the normally "disperse" diploid grew flocculently, indicating that caution must be exercised in evaluating this character. Pending further study, this character is included in the analyses only tentatively.

To determine whether phenotypically similar mutants were genetically allelic, the following procedure was used: two mutants of opposite mating type, deficient for the same metabolite, were mixed in broth to permit copulation[13] and incubated at 25°C. for 72 hours. The cells were then centrifuged, washed and plated on synthetic agar[21] lacking the metabolite for which the common requirement existed. If growth of the diploid cells occurred in 48 hours at 30°C., the mutants were considered as being non-allelic. If there were no growth, then they were taken to be allelic. The test was confirmed genetically in an analysis of two uracilless mutants ($Ur_1^-$ and $Ur_2^-$). The results are shown in table 2. The fact that uracil-independent ascospores are segregated out proves that the genes are non-allelic. The analysis in the last column in table 2 is based on the theoretical genetic segregations shown in table 3.

By back-crossing the haploid segregants to the two parental uracilless types, and carrying the diploids through a prototroph-recovery procedure, it was possible to assign the appropriate genetic constitution in those cases where it was not readily apparent by inspection. The reasoning behind the genotype-phenotype relation in table 3 is as follows: if $Ur_1$ and $Ur_2$ are two separate genes concerned in the biosynthesis of uracil, and falling either in a single biosynthetic chain of reactions, or in two separate series neither of which alone can support growth, then assuming the non-function of either gene in a haploid, it should be unable to grow without an appropriate exogenous metabolite. The genetic tests which were made with the marked strains confirmed the basic reasoning, in that heterozygosity could be demonstrated where theory demanded it should exist. It should also be noted in table 2 that mating type, and the requirements for thiamin (Th) and methionine ($Me_1$) segregated regularly.

The two uracilless mutants were tested by the auxanogram technique for their responses to various compounds. Both $Ur_1^-$ and $Ur_2^-$ were found to be capable of using uridine, but neither could use cytidilic or guanylic acids. However, $Ur_1^-$ responded to thymine, while $Ur_2^-$ did not. Loring and Pierce[21] have reported a pyrimidine mutant of *Neurospora* which was able to utilize either thymine or uracil, but which could also utilize cytidilic acid. Both of our uracilless yeast mutants appear to differ from these *Neurospora* mutants.

The prototroph-allelism test was also applied to the methionineless and adenineless mutants. At least two classes of methionineless and 5 classes of adenineless mutants were uncovered. The methionineless mutants are unable to utilize the *D*-isomer, in contrast to reported results in *E. coli.*[22] The methionineless yeasts are able to use cysteine partially, but cannot use cystine (since in the original screening procedure, cystine was present but could not replace the methionine requirement). All of the adenineless mutants can use hypoxanthine as well as adenine. They grow to varying

TABLE 4

SEGREGATION OF CHARACTERS IN THE ASCOSPORES FROM THE CROSS $a\ D^+Th^-Tr^-Ur_1^-Me_1^+ \times \alpha\ D^-Th^+Tr^+Ur_1^+Me_1^-$

| SPORES → | GENETIC CONSTITUTION OF THE FOUR ASCOSPORES | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | | | | | | 2 | | | | | | 3 | | | | | | 4 | | | | | |
| ASCUS | MT | D | Th | Tr | Ur$_1$ | Me$_1$ | MT | D | Th | Tr | Ur$_1$ | Me$_1$ | MT | D | Th | Tr | Ur$_1$ | Me$_1$ | MT | D | Th | Tr | Ur$_1$ | Me |
| A5 | α | − | − | + | − | − | a | − | + | − | − | − | a | + | + | − | + | + | | | | | | |
| A7 | a | − | − | + | + | − | α | − | − | − | − | + | α | + | + | − | + | + | | | | | | |
| A18 | α | − | + | − | + | + | a | + | − | − | − | − | a | + | − | + | + | + | | | | | | |
| A19 | α | − | + | − | + | − | α | + | − | + | − | − | a | + | − | − | − | + | a | − | + | + | + | + |
| A23 | α | − | + | + | − | + | a | − | + | − | − | − | α | + | − | − | + | + | a | + | − | + | + | − |
| A24 | α | − | − | + | − | + | a | − | + | − | − | − | α | + | − | + | + | − | a | + | + | − | + | + |
| A25 | a | − | − | − | + | − | α | + | + | + | − | − | α | + | + | − | + | + | a | − | − | + | − | + |
| A26 | a | + | + | + | + | − | a | − | + | + | + | − | α | − | − | − | − | + | α | + | − | − | − | + |
| A27 | a | + | − | + | − | + | α | − | + | − | + | + | α | − | − | − | − | − | a | + | + | + | + | − |
| A28 | α | + | − | + | + | − | α | + | + | − | − | − | a | − | + | + | + | + | | | | | | |
| A29 | a | − | + | − | − | + | α | − | − | − | − | − | α | + | − | + | + | − | | | | | | |
| A32 | a | + | + | + | − | − | α | − | − | − | − | +? | α | − | − | − | + | + | a | + | + | + | + | + |
| A33 | a | − | − | + | − | − | α | + | − | + | − | + | α | − | + | − | + | + | | | | | | |
| A34 | a | + | − | − | + | − | α | − | + | + | − | − | a | − | + | + | + | + | α | + | − | − | − | + |
| A35 | a | + | − | − | + | − | α | − | + | + | + | + | α | + | − | + | − | − | | | | | | |
| A37 | α | − | − | + | − | − | α | − | + | − | − | + | a | + | + | − | + | + | a | + | − | + | + | − |
| A38 | a | − | − | + | − | − | a | − | − | + | − | + | α | + | + | − | + | + | α | + | + | − | + | − |
| A39 | α | + | − | + | − | + | a | + | + | − | + | + | α | − | − | − | + | − | a | − | + | + | − | − |
| A40 | α | + | + | − | + | + | a | − | + | + | − | − | a | − | − | + | − | + | | | | | | |
| A41 | a | − | − | + | − | + | α | − | − | − | − | − | a | + | + | + | + | + | | | | | | |
| A42 | a | + | − | + | + | + | a | + | − | − | + | + | α | − | + | − | − | − | α | − | + | + | − | − |
| A43 | a | + | + | + | − | + | a | + | − | − | + | + | α | − | + | + | + | − | α | − | − | − | − | − |
| A44 | a | − | − | + | + | − | α | + | + | + | − | + | α | + | + | − | − | − | a | − | − | − | + | + |
| A45 | α | + | − | + | + | − | α | + | + | − | − | − | a | − | − | − | + | + | a | − | + | + | − | + |
| A46 | a | − | − | − | + | − | a | + | + | + | + | + | α | − | − | + | − | − | | | | | | |
| A47 | α | − | + | − | + | − | α | + | − | − | + | + | a | − | + | + | − | − | a | + | − | + | − | + |
| A48 | a | + | − | − | + | − | α | + | − | + | − | + | a | − | + | + | + | − | α | − | + | − | − | + |

extents on adenosine, but only one can use adenylic acid and adenosinetriphosphate ($\alpha$, adenineless, pink strain of Reaume and Tatum[5]). Work is in progress to characterize these mutants further.

To illustrate the genetic results obtained, the analysis of the cross $a\ D^+Th^-Tr^-Ur_1^-Me_1^+ \times \alpha\ D^-Th^+Tr^+Ur_1^+Me_1^-$ is presented in table 4. Other characteristics included in this cross were growth rate and drug resistance, but these do not appear to segregate as single factors, and will be discussed elsewhere. As is readily apparent from the data in table 4, the clones derived from single ascospores showed considerable regularity (there being only one questionable result). Mendelian principles of segregation and independent assortment certainly appear applicable to this material. $X^2$ (Chi-square) analysis was carried out for all possible combinations of genes, and there was evidence of linkage between two of the characters, $D$ and $Ur_1$ ($X^2 = 7.2$, one degree of freedom, $p = 0.007$). The range of $X^2$ values for the other characters was from $X^2 = 0$ (MT:D, D:Tr, and Th:$Me_1$) to $X^2 = 3.7$ (MT:Tr). Of outstanding significance is the complete confirmation of Lindegren's observation[12] of two stable mating types in this strain of *S. cerevisiae*. Although the strength of the mating reaction varies with each clone, as reflected by the frequency of appearance of conjugating cells in microscopic examination of crosses of segregants × tester stocks, the heterothallic status of this strain seems clear. These results, and those obtained with other crosses, indicate that, for the characters studied, inheritance is Mendelian and completely analogous to that in higher organisms.

*Discussion.*—The problem posed at the outset of this work was to devise means whereby the inheritance of characters controlled by single genes could be studied under controlled conditions. The stability of the haploid mating types[12] allows considerable latitude in the manipulations to which they can be subjected, and still retain their mating ability. Thus, it was relatively simple to produce stable mutants which possessed the power to conjugate with mutants of the opposite mating type. The prototroph-recovery procedure insures the isolation of diploids arising from heterothallic conjugation. In addition, it is a very effective procedure in recovering diploids where conjugation is poor. Thus, for example, if only 2 out of $10^6$ cells were able to conjugate, the diploid heterozygote could readily be recovered from a heavily seeded plate of minimal agar. The dispatch and sureness with which this procedure also screens for allelism is of great potential importance in studies of biochemical genetics with this organism. The ability of *S. cerevisiae* to survive either as a haploid or a diploid makes it a very flexible tool for different types of genetic investigations.

It seems reasonable to conclude that the characters studied in these crosses segregated according to Mendelian principles. One cannot state that all of the genes of this yeast segregate regularly, although it seems not unlikely that the same basic pattern should be found in most cases where single genes are being studied. With regard to the inheritance of adaptive enzymes, little can be said. In order to decide whether the mechanism of inheritance is through cytoplasmic genetic units as suggested by Lindegren and Lindegren[1] and Spiegelman,[2] or by genes controlling rapid and slow fermentation, as reported by Winge and Roberts,[14] requires further experimental evidence. It would appear worth while to determine how many processes are actually being measured in the course of adaptation, and then to try to follow the inheritance of a single reaction instead of an entire complex, for the purpose of clarifying the inheritance of this type of reaction.

*Summary.*—Biochemically deficient mutants of *Saccharomyces cerevisiae* Hansen were obtained after ultra-violet irradiation of haploid mating strains. Various mutants were crossed, and heterozygous diploids obtained by a prototroph-recovery procedure. The same procedure was used in studying allelism between phenotypically similar mutants. Genetic analyses indicated that the characters studied segregated independently and regularly according to Mendelian principles.

---

[*] This material was abstracted from a dissertation, presented by the senior author for the degree of Doctor of Philosophy in Yale University.

[†] This work was done while the senior author was the recipient of a Standard Brands, Inc., Fellowship in Microbiology at Yale.

[‡] The symbols used to represent genetic characters are as follows: $a$ and $\alpha$ = mating type (MT); Ur = uracil; Me = methionine; and Ad = adenine. Superscripts $^+$ and $^-$ refer to nutritional status ($^+$ = independent, $^-$ = dependent), and independence $^{(+)}$ is dominant over dependence $^{(-)}$. Subscripts refer to different alleles with the same apparent phenotypic expression. Thus, $a\,Th^-Tr^-Ur_1^-Me_1^+ \times \alpha\,Th^+Tr^+Ur^{1+}Me_1^-$ refers to a cross between $a$ thiaminless, tryptophanless, uracilless (locus 1), and $\alpha$ methionineless (locus 1). + and − in the body of a table refer to genetic constitution, i.e., independence and dependence, respectively.

[1] Lindegren, C. C., and Lindegren, G., *Cold Spring Harbor Symposia, Quant. Biol.*, **11**, 115–129 (1946).

[2] Spiegelman, S., *Ibid.*, **11**, 256–277 (1946).

[3] Lindegren, C. C., and Lindegren, G., these PROCEEDINGS, **33**, 314–318 (1947).

[4] Beadle, G. W., and Tatum, E. L., *Ibid.*, **27**, 499–506 (1941).

[5] Reaume, S. E., and Tatum, E. L., in press (1949).

[6] Beadle, G. W. *Chem. Rev.*, **37**, 15–96 (1945).

[7] Fries, N., *Arkiv. für Botanik*, **32A**, 1–9 (1945).

[8] Lederberg, J., *Genetics*, **32**, 505–525 (1947).

[9] Lederberg, J., and Tatum, E. L., *J. Biol. Chem.*, **165**, 381–382 (1946).

[10] Meyersburg, R., Pomper, S., and Cutter, V. M., Jr., *Science*, **109**, 446–447 (1949).

[11] Po[illegible]S., Abst. in *Am. J. Botany*, **35**, 802 (1948).

[12] Lin[illegible], C. C., and Lindegren, G., *Ann. Missouri Botan. Garden*, **30**, 453–469 (1943).

[13] Lindegren, C. C., and Lindegren, G., these PROCEEDINGS, **29**, 306–308 (1943).

[14] Winge, Ö., and Roberts, C., *Compt. rend. Labor. Carlsberg, ser. physiol.*, **24**, 263–315 (1948).

[15] Ryan, F. J., and Lederberg, J., these PROCEEDINGS, **32**, 163–173 (1946).

[16] Winge, Ö., and Laustsen, O., *Compt. rend. Labor. Carlsberg, ser. physiol.*, **22**, 99–116 (1937).

[17] Lindegren, C. C., and Lindegren, G., *Botan. Gaz.*, **105**, 304–316 (1944).

[18] Graham, V. R., and Hastings, E. G., *Can. J. Research*, **19**, 251–256 (1941).

[19] Stantial, H., *Trans. Roy. Soc. Can.*, 3rd ser., III, **29**, 175–188 (1935).

[20] Fitz, G. W., *Science*, **79**, 233–234 (1934).

[21] Loring, H. S., and Pierce, J. G., *J. Biol. Chem.*, **153**, 61–69 (1944).

[22] Simmonds, S., *Ibid.*, **174**, 717–722 (1948).

---

# *VARIATION IN THE AMOUNT OF DESOXYRIBOSE NUCLEIC ACID IN DIFFERENT TISSUES OF TRADESCANTIA*

BY FRANZ SCHRADER AND CECILIE LEUCHTENBERGER

DEPARTMENT OF ZOOLOGY, COLUMBIA UNIVERSITY

Communicated by L. C. Dunn, June 3, 1949

It has become a text book generalization that all cells of an organism have the same chromosomal constitution. To the cytogeneticist such a generalization offers little hope for a solution of the great problem of differentiation, though the outlook is not quite so discouraging on the chemical level. The biochemical analyses of Chargaff and others[3, 13] have established that the composition of desoxyribose nucleic acid (DNA), which forms so important a constituent of the chromosome, is not always the same and may thus furnish a basis for variation. Further, it has recently been shown (Pollister and Leuchtenberger,[9] and Leuchtenberger[7]) that physical differences resting on various degrees of polymerization may underlie further variations of DNA, and therewith some very important steps in the analysis of differentiation have been initiated.

Indeed, the cytological investigations of Geitler and a few other workers have furnished evidence that the initial generalization was altogether too sweeping in its claims. In many organisms there are tissues whose cells have higher numbers of chromosomes than are encountered in the diploid complement. The recent work of Geitler and of Huskins has made it clear that polyploidy or polysomaty is characteristic of many fully differentiated tissues, and, as Huskins points out, such "data on endopolyploidy permit us to assume quantitatively different gene action in different tissues" (Huskins).[5]

But, though the cytologist was prone to accept an identity of chromosomal number for all normal cells in an organism prior to these discoveries, he could not possibly escape the fact that the chromosomes in the cleaving egg are often larger than those of the blastula and gastrula, and that in later stages the chromosomes of some tissues are uniformly larger than those of others. Here is a quantitative difference which, oddly enough, was rarely considered. Possibly this was due to a recognition of the possibility that in such instances the larger chromosomes have their materials in a more diluted condition than the small ones, and that their essential constituents therefore did not differ in amount. To be sure, visual comparison under the microscope seems, in at least some cases, to speak against such an explanation, but it may well be recognized that there is involved such a multiplicity of factors in the optical and staining properties of chromosomes that a final decision would be very difficult to obtain.

The recent findings of Boivin, Vendrely and Vendrely,[2] Vendrely and Vendrely[12] and Mirsky and Ris,[8] would seem to suggest that such caution is indeed warranted. Working on tissues in vertebrates, they arrived at the conclusion that all the diploid cells of an organism carry identical amounts of DNA. Their conclusions were based on a gross biochemical analysis of nuclei separated from the cytoplasmic constituents of the cell. In evaluating these results it should be noted that the amount of DNA determined for the nuclei of certain tissues by the Vendrelys differs from the amount reported for the same tissues by Mirsky and Ris. Indeed, in their values for bull sperms the foregoing authors differ not only from each other but also from Zittle and O'Dell.[14] Any attempt to generalize on these findings must therefore await the results of further analyses.

The question at issue is so important that an investigation of completely different material seemed warranted. For this purpose the plant Tradescantia paludosa was used, all material coming from plants of the same clone.[1] The tissues tested were root tips, the petals of small, unopened buds, the basal parts of leaves and certain meiotic stages. All were fixed in Carnoy's acetic alcohol (1:3) and sectioned at 15 $\mu$. Staining was always done in exactly the same way under standardized conditions and to further assure an identity of treatment, sections of root and bud tissue and, in other cases, bud and leaf tissue were mounted on the same slides. Both the root and bud tissues studied showed only metaphases with the diploid number of 12 chromosomes; in the leaf tissue almost no mitoses were encountered.

The Feulgen reaction, which is known to have a high specificity for DNA (Stowell,[11] DiStefano[4]), was used throughout. Its reproducibility has recently been shown by Leuchtenberger in animal material. In order to measure the amounts of Feulgen dye in individual nuclei, we used the photometric, microscopic method of Pollister and Ris[10] which permits the estimation of relative amounts of colored precipitates in the nuclei of fixed and stained sections. The Feulgen reaction was measured in terms of the absorption of the green light isolated from a zirconium lamp by a Farrand interference filter with a peak transmission in parallel light at 557 millimicrons. The photo tube was an R. C. A. Photomultiplier, Type 931 A. The mean extinctions (mean of 10–20 nuclei) obtained appear in the Table as $E_{550}$. For a better evaluation of the absorption data, extinction values were considered in terms of nuclear sizes and this was done by using $E_{550} \times \pi r^2$ as an arbitrary unit (see Column 6 in table 1). For the absorption measurements of the Feulgen dye in root, bud and leaf tissues, nuclei of the same mean nuclear diameter (8.6 microns) were selected; similarly, comparisons between prophases in root tips and meiotic pachytene stages were made in nuclei with a mean diameter of 10.8 microns. The DNA per nucleus was computed as described by DiStefano and, since such computations involve several assumptions and are subject to some errors (Leuchtenberger[7]), it must be considered that these DNA values are relative and not absolute.

TABLE 1

COMPARISON OF DESOXYRIBOSENUCLEIC ACID (FEULGEN) IN DIFFERENT TISSUES OF TRADESCANTIA PALUDOSA

| STAGE | TYPE OF TISSUE | MEAN NUCLEAR DIAMETER IN MICRONS | NUMBER OF NUCLEI MEASURED | $E_{550}$ | $E_{550} \times \pi r^2$ | RELATIVE AMOUNT OF DNA IN $10^{-9}$ MG.* PER NUCLEUS |
|---|---|---|---|---|---|---|
| Interphase | Root tip | 8.6 | 20 | 0.443 ± 0.014 | 25.7 ± 0.78 | 5.5 |
| Early prophase | Root tip | 10.8 | 10 | 0.570 ± 0.025 | 51.4 ± 2.40 | 11.0 |
| Interphase | Leaf | 8.6 | 17 | 0.735 ± 0.02 | 42.7 ± 1.29 | 9.0 |
| Interphase | Bud tapetum | 8.6 | 20 | 0.969 ± 0.02 | 56.3 ± 1.68 | 12.0 |
| Meiosis pachytene | Bud anther | 10.8 | 10 | 0.939 ± 0.02 | 84.7 ± 2.4 | 18.0 |
| Meiosis microspore interphase | Bud anther | 7.0 | 10 | 0.733 ± 0.08 | 25.7 ± 2.9 | 6.0 |

* Values in these columns involve several assumptions and therefore are only relative.

On the basis of measurements such as are tabulated here, it appears that interphase nuclei of different somatic tissues of Tradescantia (root, bud and leaf) do not carry identical amounts of DNA even though, in bud and root tissue at least, the same (diploid) chromosome number is involved. So large are the differences involved that there would seem to be no escape from this conclusion.

It should further be noted that pachytene nuclei of meiotic cells show considerably more DNA than equally sized somatic prophase nuclei. Also, such pachytene nuclei carry three times as much DNA as the interphase nuclei after the second division (microspores). It is likely that the actual ratio is the theoretically expected 4:1 and that the values obtained for the pachytene nuclei appear low because of the uneven distribution of chromosome material at that stage. Especially striking is the fact that the microspore carries an amount of DNA corresponding very closely to that of the root tip nucleus. This is a coincidence which in itself demonstrates that haploid nuclei resulting from two meiotic divisions do not necessarily carry only half the DNA contained in diploid, somatic cells.

It is of interest that if Tradescantia tissues are fixed in 20% Formalin, absorption measurements of Fuelgen-stained nuclei give extinction values about twice as great as after Carnoy fixation. In spite of this increase the ratios of relative amounts of DNA in different tissues (for instance, between bud and root) are essentially the same for both types of fixation.

The unavoidable conclusion is that the amount of DNA carried in a given chromosome may vary in different tissues. The explanation of such variation may involve one of two basic conceptions. Thus, it is possible that the DNA has no exact quantitative relation to the basic structure of the chromosome and may vary according to the age or other physiological conditions of the nucleus or its surroundings. An interpretation of this sort has been made by several cytologists and is implied in such terms as "nucleic acid starvation" and "supercharging of nucleic acid." As long as the structural and chemical composition of the chromosome is not better known, such a possibility must be considered. Another interpretation rests on the cytological concept that each chromosome is composed of a number of threads or chromonemata. If each of such threads is associated with a definite and identical quantity of DNA, then the total amount of DNA in a chromosome depends on the number of chromonemata of which it is composed. A chromosome with a large number of chromonemata—that is, of a high degree of polyteny—would therefore contain more DNA than an otherwise homologous chromosome of a lower degree of polyteny.

On the whole, this second interpretation is at present the most acceptable, if only because we already know definitely that polyteny exists. On such a basis the haploid chromosome complement of the Tradescantia microspore carries practically as much DNA as the diploid root tip because each of its chromosomes is composed of twice as many chromonemata. It is a strong argument for the claim that reorientation in some of our cytogenetic concepts is called for—especially with respect to the numerical relation of the chromonemata to the chromosomes.

The fact that the different amounts of DNA in Tradescantia do not show any obvious multiple relation to each other may be due to several causes. First of all is the possibility that, for mechanical reasons, the photometric determinations are still subject to a certain degree of error, as we have already stated. But it is also conceivable that, as Huskins and Steinitz[6] have pointed out in their analysis of various types of endomitosis in Rhoeo, certain chromosomes may divide prior to others of the same complement. This may also obtain in the internal replication involved in polyteny, with the result that intermediate values are obtained for some tissues.

If DNA has the exact quantitative relation to the individual chromonema that has been considered in the foregoing, the recent conclusions of Boivin and the Vendrelys are not surprising. They mean only that the vertebrate tissues analyzed all have nuclei in which the same degree of polyteny obtains. Such data do not necessarily give final information on the number of chromosomes in such tissues for, as our findings in Tradescantia show, the quantity of DNA contained in a nucleus is probably dependent on the grade of polyteny in its chromosomes.

*Summary.*—Contrary to the findings of several workers for the tissues of vertebrates, the nuclei in different tissues of the plant Tradescantia carry different amounts of DNA. This is not attributable to various degrees of polyploidy in this case but may be due to different degrees of polyteny that obtain in the various tissues.

---

[1] We are greatly indebted to Professor C. L. Huskins, of the University of Wisconsin, who selected and fixed the various types of tissue that we required.

[2] Boivin, A., Vendrely, R., and Vendrely, C., *C. R. Acad. Sci.*, Paris, **226**, 106 (1948).

[3] Chargaff, E., and Saidel, H. F., *J. Biol. Chem.*, **177**, 417–428 (1949).

[4] DiStefano, H. S., *Chromosoma*, **3**, 282–292 (1948).

[5] Huskins, C. L., *Am. Nat.*, **81**, 401–434 (1947).

[6] Huskins, C. L., and Steinitz, L., *J. Hered.*, **39**, 34–43 (1948).

[7] Leuchtenberger, C., Chromosoma in press (1949).

[8] Mirsky, A. E., and Ris, H., *Nature*, **163**, 666–667 (1949).

[9] Pollister, A. W., and Leuchtenberger, C., PROC. NATL. ACAD. SCI., **35** 111–116 (1949).

[10] Pollister, A. W., and Ris, H., *Cold Spring Harbor Symp.*, XII 147–157 (1947).

[11] Stowell, R. E., *Stain Tech.*, **20**, 45–48 (1945).

[12] Vendrely, R., and Vendrely, C., *Experientia*, **4**, 434–436 (1948).

[13] Vischer, E., Zamenhoff, S., and Chargaff, E., *J. Biol. Chem.*, **177**, 429–438 (1949).

[14] Zittle, C. A., and O'Dell, R. A., *J. Biol. Chem.*, **140**, 899–907 (1941).

---

## *CHROMOSOMES OF A FOX HYBRID* (*Alopex-Vulpes*)*

BY LOUISE WIPF† AND RICHARD M. SHACKELFORD‡

DEPARTMENT OF GENETICS, UNIVERSITY OF WISCONSIN

Communicated by C. E. Allen, June 1, 1949

Numerous attempts have been made to obtain offspring by mating animals classified as belonging to different species or even different genera. When young have been the result of such matings, they often prove to be semisterile or completely sterile in matings *inter se* or back to either of the parental groups. The mule has long been regarded as the classical example of an intergeneric hybrid which falls into this category. From time to time since the red fox (*Vulpes vulpes* L.-*V. fulva* species) and the arctic fox (*Alopex lagopus* L.) have been brought under domestication, breeders have produced hybrids from the two species with the idea that, by back-crossing, the size of the arctics could be increased and their higher productivity introduced into the color phases of the red fox. This objective has not met with success to date, however, apparently as a result of the complete sterility of this intergeneric hybrid.

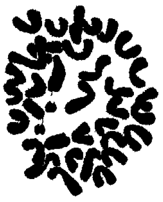
Cole and Shackelford[1] have described *Alopex-Vulpes* hybrids and discussed their breeding behavior in detail. One of the males from the group designated as Case 8 by these workers furnished material for the chromosome studies reported in this paper. Tissues from the black (silver) color phase of the red fox came from the University of Wisconsin ranch and those from the arctic fox were collected through the courtesy of Messrs. John A. Fromm and Loyal Wells of Fromm Bros., Nieman and Co., Inc., Thiensville, Wis.

Individuals of the two parental species and the hybrid were castrated during the breeding season (1945) in order to obtain meiotic as well as mitotic divisions. Although the breeding seasons of the two species overlap, that of the arctic fox averages a later date than that of the red fox in this locality. Consequently, the testes from the arctic fox were taken March 21, and those from the red fox February 10. It was assumed that the height of the breeding season for the hybrid was probably between those for the parent species, so one testis was removed March 7, 1945. Studies of this material suggested that the testis of the hybrid might have been removed too early in the breeding season, so the next year (1946) the remaining testis was taken at approximately the date found satisfactory for the arctic fox, namely March 19.

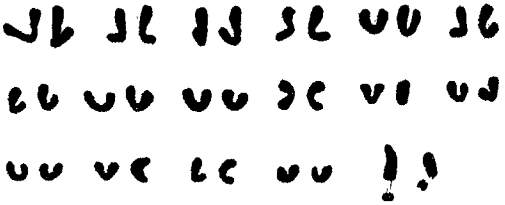
Seminiferous tubules were teased from the testes directly into Carnoy's alcohol-acetic acid-chloroform fixative (7:2:1). An hour later the material was transferred to 80% alcohol for storage. Temporary aceto-carmine smears were made and the desired division figures observed.

Andres[2] and Wipf and Shackelford[3] reported 34 as the diploid chromosome number for *Vulpes*. The material of the red fox used in this study came from a different animal from any of those used in the 1942 report,[3] and confirmed 34 as the chromosome number for *V. vulpes* (figure 1 *A*). The present study is also in agreement with Andres'[2] count of 52 for *Alopex* (figure 3 *A*). Wodsedalek[4] gave 42 as the number for the red fox, Bishop[5] 32 and Makino[6] 38.

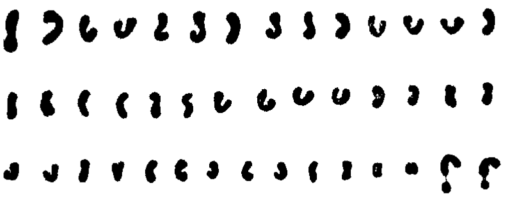
Considerable differences can be noted in size and shape between the chromosomes of the two parent species at mitosis; the chromosomes of the red fox are much larger and longer, the largest being more than twice the length of the largest in the arctic. Both species show satellite attachments in one pair. No sex chromosomes can be distinguished on the basis of heteropycnosis, differences in size or other morphological characters. Within each species it is relatively easy to pair a chromosome with another which appears to be its homologue (figures 1 *B* and 3 *B*). At the first reduction division the chromosomes of *Vulpes* (figure 4) are larger than



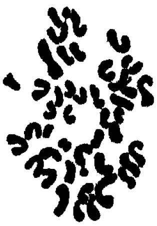
1A



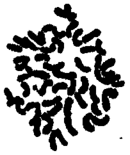
1B



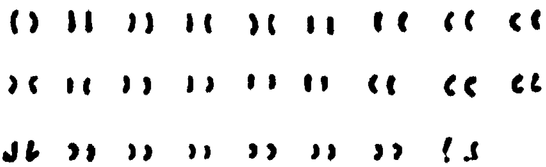
2B



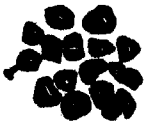
2A



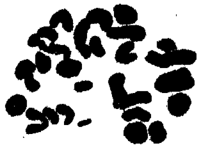
3A



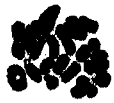
3B



4



5



6

FIGURE 1 (*A*)

*Vulpes* chromosomes on a mitotic equatorial plate. $2n = 34$.

FIGURE 1 (*B*)

Chromosomes from figure 1 (*A*) arranged as homologous pairs.

FIGURE 2 (*A*)

Mitotic equatorial plate from an *Alopex-Vulpes* hybrid showing a chromosome number of 43.

*Captions continued on facing page.* →

those of *Alopex* (figure 6) as was the case in mitotic divisions, and chromosomes that appear morphologically similar are paired in both species.

Polar views of the equatorial plate stage of mitotic figures from the *Alopex-Vulpes* hybrid show 43 chromosomes (figure 2 *A*). Some of the chromosomes are large and can readily be recognized as coming from the *Vulpes* parent; others are smaller and more closely resemble those of *Alopex* (figure 2 *B*). At the first reduction division a mixture of paired and single chromosomes is found on the equatorial plate (figure 5). Side views of equatorial plates, anaphases and telophases show no abnormalities such as precocity or lagging in either mitotic or meiotic divisions.

It is doubtful that mature sperms are formed in the hybrid. The cell type which comes nearest resembling a sperm is large and irregular in outline and lacks the compact appearance of *Alopex* and *Vulpes* sperm heads. Whatever may be the causative element, it is possible that the sterility of the *Alopex-Vulpes* hybrid, at least in the male, stems from the inability to produce mature gametes.

Other workers have reported irregularities of various kinds in the formation of gametes in a number of interspecific crosses in mammals. Wodsedalek[7] noted that in the mule, chromosomes were often scattered on the spindle as well as in the cytoplasm at the metaphase stage and multiple and distorted spindles were often seen. Dice[8], working on dog-coyote hybrids, reported that "the maturation divisions seem to be highly abnormal." He found no divisions at any stage later than early anaphase of the first reduction division. In cattle-bison hybrids (Deakin, Muir and Smith[9]), the seminiferous tubules were found to be in a degenerative state with no germinal epithelium. There were indications of active mitosis and a degree of differentiation of spermatids toward spermatozoa in some of the tubules.

* Joint contribution from the Wisconsin Conservation Department's Experimental Game and Fur Farm, Poynette, and the Department of Genetics (Paper No. 401), Agricultural Experiment Station, University of Wisconsin. Published with the approval of the Director of the Station.

† Genetics Department, University of Wisconsin.

---

FIGURE 2 (*B*)

Chromosomes from figure 2 (*A*) placed in a linear arrangement.

FIGURE 3 (*A*)

*Alopex* chromosomes on a mitotic equatorial plate. $2n = 52$.

FIGURE 3 (*B*)

Chromosomes from figure 3 (*A*) arranged as homologous pairs.

FIGURES 4–6

Polar views of equatorial plates at the first meiotic division. Figure 4, *Vulpes;* figure 5, *Alopex-Vulpes* hybrid; figure 6, *Alopex*. All figures × 1800.

‡ Genetics Department, University of Wisconsin, and Agent, Bureau of Animal Industry, U.S.D.A.

[1] Cole, L. J., and Shackelford, R. M., *Trans. Wis. Acad. Sci.*, **38,** 315–332 (1946).

[2] Andres, A. H., *Cytologia*, **9,** 35–37 (1938).

[3] Wipf, L., and Shackelford, R. M., PROC. NATL. ACAD. SCI., **28,** 265–268 (1942).

[4] Wodsedalek, J. E., *Anat. Rec.*, **51,** 70 (1931).

[5] Bishop, D. W., *Ibid.*, **84,** 99–115 (1942).

[6] Makino, Sajiro, *J. Faculty Sci., Hokkaido Imp. Univ., Ser. VI*, **9** (4), 345–357 (1947).

[7] Wodsedalek, J. E., *Biol. Bull.*, **30,** 1–56 (1916).

[8] Dice, L. R., *J. Mammal.*, **23,** 186–192 (1942).

[9] Deakin, A., Muir, G. W., and Smith, A. G., Publ. 479 (Tech. Bull. 2), Dom. Canada, Dept. of Agric., 30 pp. (1935).

---

# *CONTINUED FRACTIONS, ALGEBRAIC FUNCTIONS AND THE PADÉ TABLE*

BY RICHARD BELLMAN AND ERNST G. STRAUS

DEPARTMENT OF MATHEMATICS, STANFORD UNIVERSITY, AND DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA AT LOS ANGELES

Communicated by S. Lefschetz, June 22, 1949

*1. Introduction.*—Given a function $f(x)$ possessing a power series development in the neighborhood of the origin, the problem, first posed by Padé, which we wish to consider is that of determining the polynomials $P_{m,n}$; $Q_{m,n}$ the first of degree $m$, the second of degree $n$, having the property that

$$P_{m,n}(x) + f(x)Q_{m,n}(x) = x^{m+n+1} + \ldots. \tag{1}$$

The sets $(P_{n,n}; Q_{n,n})$, $(P_{n,n-1}; Q_{n,n-1})$ are obtained from the continued fraction expansion

$$f(x) = \sum_{n=0}^{\infty} a_n x^n \sim b_0 + \frac{b_1 x}{1+}\,\frac{b_2 x}{1+}\ldots, \tag{2}$$

associated with $f(x)$, cf. Perron.[1]

In general, the problem posed above is difficult despite the fact that the coefficients of the unknown polynomials $P_{m,n}$; $Q_{m,n}$ may be obtained formally by solving the $m + n + 1$ linear equations equivalent to (1). In this way necessary and sufficient conditions may be obtained for the existence of a unique solution to (1). However, this mode of procedure is unsatisfactory in many respects, since it is virtually impossible to apply the criteria found in this way, and the structure of the coefficients $b_n$, and the polynomials $P_{m,n}$; $Q_{m,n}$ remains hidden.

For these reasons, various artifices have been employed to find the continued fraction expansion of $f(x)$ and the elements $P_{m,n}$; $Q_{m,n}$ of the Padé table. These depend, naturally, upon the particular functional properties of the function under consideration. Thus there are techniques available for functions associated with the hyper-geometric function, for functions satisfying linear differential equations with polynomial coefficients, for functions of the form $\int_{-\infty}^{\infty} d\alpha(t)/(x - t)$ and so on, cf. Perron.[2]

The primary purpose of this paper is to present a method applicable, in principle, to the class of all algebraic functions; in practice, difficulties of various types occur. Nevertheless, using this method we are able to obtain many new continued fraction expansions quite easily. Some examples will be given below.

We also present a method applicable only to algebraic functions defined by equations of the form

$$y^n + p_1(x)y^{n-1} + \ldots + p_n(x) = 0 \tag{3}$$

where the $p_i(x)$ are linear polynomials in $x$. The most interesting functions of this class are those satisfying the simple equations, $y^n - (1 + x) = 0$.

Let us now sketch the first method. Turning to (1), we note that if $f(x)$ is algebraic, so also is the left-hand side of (1). According to the right-hand side of the equation, this algebraic function must have a zero of order $m + n + 1$ at $x = 0$. Furthermore, the set of poles of this algebraic function is easily determined, consisting of the poles of $f(x)$, which may be found by reference to the defining equation, and the point $x = \infty$. Since an algebraic function is completely determined, apart from a multiplying constant, once its zeroes and poles are known together with their multiplicities, the expression $P_{m,n}(x) + Q_{m,n}(x)f(x)$ will be theoretically known provided that we can extract enough information from (1), the general result for all algebraic functions that the number of zeroes is equal to the number of poles, and the further relations, depending upon the genus of the defining curve, that must hold between the zeroes and the poles.

In some cases, a count of the zeroes and poles is sufficient to fix the left-hand side of (1). In order to furnish an explicit expression, we introduce the transcendental functions which uniformize the curve, e.g., for curves of genus zero, the trigonometric functions, for curves of genus one, the elliptic functions, and so on. One of the practical difficulties is that the transcendental functions, required in general for the uniformization of curves of genus greater than or equal to two, are not readily accessible.

If the initial argument on the zeroes and poles does not suffice, we must utilize the specific form of the left-hand side of (1). An example is given below.

Proofs of the results stated here, and further results, will be published separately.

*2. Examples of the First Method.*—Consider the curve defined by $y^2 = ax^3 + bx^2 + cx + d$, $a \neq 0$. Let us impose the condition that $ax^3 + bx^2 + cx + d = 0$ have no double root. In this case, the equation of the curve may be written in the form

$$y^2 = 4(c_1x + c_2)^3 - g_2(c_1x + c_2) - g_3, \qquad g_2^3 \neq 27g_3^2, \qquad c_1 \neq 0, \tag{1}$$

and has the parametrization,

$$c_1x + c_2 = p(u), \qquad y = P'(u) \tag{2}$$

where $P(u)$ is the Weierstrassian elliptic function. Let $2w_1$, $2w_2$ be a set of primitive periods of $P(u)$, and $\alpha$, $-\alpha$ be the roots of $p(u) - c_2 = 0$. Then

THEOREM. *The necessary and sufficient condition for the existence of a unique sequence of polynomials* $\{P_n, Q_{n-1}\}$, $P_n$ *of degree* $n$, $Q_{n-1}$ *of degree* $n - 1$, *satisfying*

$$P_n(x) + \sqrt{ax^3 + bx^2 + cx + d}\, Q_{n-1}(x) = x^{2n} + \ldots \tag{3}$$

*is that*

$$\alpha \neq \frac{m_1w_1 + m_2w_2}{n}, \qquad m_1, m_2 = 0, \pm 1, \pm 2, \ldots; \qquad n = 1, 2, \ldots. \tag{4}$$

*If this condition is satisfied, we have*

$$P_n\left(\frac{p(u) - c_2}{c_1}\right) + p'(u)Q_{n-1}\left(\frac{p(u) - c_2}{c_1}\right) = \frac{c_1^{-2n}p'(\alpha)^{2n}\sigma(\alpha)^{2n+1}\sigma(u + 2n\alpha)\sigma(u - \alpha)^{2n}}{\sigma((2n + 1)\alpha)\sigma(u)^{2n+1}}. \tag{5}$$

*From this relation, the elements of the continued fraction expansion for* $\sqrt{ax^3 + bx^2 + cx + d}$ *may be derived.*

*Sketch of Proof:* If we introduce the parametrization (2) then equation (3) becomes

$$P_n\left(\frac{p(u) - c_2}{c_1}\right) + p'(u)Q_{n-1}\left(\frac{p(u) - c_2}{c_1}\right) = \left(\frac{p(u) - c_2}{c_1}\right)^{2n} + \ldots. \tag{6}$$

The left-hand side of the above equation is an elliptic function with a pole of order $2n + 1$ at $u = 0$, and, by virtue of the right-hand side, with a zero of order $2n$ at $u = \alpha$. Although $p(u) - c_2 = 0$ has two zeroes, at $u = \pm\alpha$, only one is a zero of the left-hand side, the other being a zero of the conjugate. Since the sum of the zero points must be congruent to the sum of the points at which the function has poles, multiple points counted multiply, we are led to the expression

$$P_n\left(\frac{p(u) - c_2}{c_1}\right) + p'(u)Q_{n-1}\left(\frac{p(u) - c_2}{c_1}\right) = d_n\sigma(u - \alpha)^{2n}\sigma(u + 2n\alpha)/\sigma(u)^{2n+1}. \quad (7)$$

The constant is now easily obtained from the normalization condition that the coefficient of $x^{2n}$ in (3) be unity.

Once the sequence $\{P_n, O_{n-1}\}$ has been obtained, an elementary result from the theory of continued fractions permits us to deduce the analogous sequence where both polynomials have the same degree, and thence the elements of the continued fraction.

As a simple illustration of the case where a count of the zeroes and poles is not sufficient, let us consider the function $y = (1 + x)\sqrt{1 - x^2}$. Using the obvious parametric representation, the problem takes the form of determining the trigonometric polynomials $P_n$, $Q_{n-1}$ such that

$$P_n(\sin\theta) + Q_{n-1}(\sin\theta)(1 + \sin\theta)\cos\theta = (\sin\theta)^{2n} + \ldots. \quad (8)$$

Here the left side of the equation is a trigonometric polynomial of degree $n + 1$, possessing, in consequence, $2n + 2$ zeroes, of which only $2n$ are determined by the equation. Therefore,

$$P_n(\sin\theta) + Q_{n-1}(\sin\theta)(1 + \sin\theta)\cos\theta = (1 - \cos\theta)^n A_1(\theta) \quad (9)$$

where $A_1(\theta)$ is a trigonometric polynomial of the first degree. From this we obtain, via the substitution $\theta \to \pi - \theta$

$$P_n(\sin\theta) = [(1 - \cos\theta)^n A_1(\theta) + (1 + \cos\theta)^n A_1(\pi - \theta)]/2,$$
$$(1 + \sin\theta)Q_{n-1}(\sin\theta) = [(1 - \cos\theta)^n A_1(\theta) - (1 + \cos\theta)^n A_1(\pi - \theta)]/2\cos\theta. \quad (10)$$

The unknown linear polynomial is now readily found.

We remark, however, that the sequence $\{P_{n,\,n-2};\ Q_{n,\,n-2}\}$ can be determined by the first method. In general, for every algebraic function there is an integer $k$ such that the sequence of polynomials $\{P_{n,\,n-k};\ Q_{n,\,n-k}\}$ is determined by the first method.

A similar, but more complicated investigation is required to derive the continued fraction associated with $y^2 = ax^4 + bx^2 + cx^3 + dx + e$. This method is applicable to the treatment of $N$-dimensional continued fractions, but lack of space does not permit any example of this here.

*3. An Example of the Second Method.*—We seek to determine the polynomials $P_{n,\,i}$, $i = 1, 2, 3$, each of degree $n$, satisfying the equation

$$P_{n,\,1} + P_{n,\,2}(1 + x)^{1/3} + P_{n,\,3}(1 + x)^{2/3} = x^{3n+2} + \ldots. \quad (1)$$

It can be shown that these polynomials are unique. Choose $A_1$, a linear polynomial, and $A_2$, $A_3$ constants, so that

$$A_1 + A_2(1 + x)^{1/3} + A_3(1 + x)^{2/3} = x^3 + \dots \qquad (2)$$

Multiplying (1) and (2), and taking account of the uniqueness of the $P_{n+1,\,i}$ we obtain the set of linear difference equations,

$$\begin{aligned} P_{n+1,\,1} &= A_1P_{n,\,1} + A_3(1 + x)P_{n,\,2} + A_2(1 + x)P_{n,\,3} \\ P_{n+1,\,2} &= A_2P_{n,\,1} + A_1P_{n,\,2} + A_3(1 + x)P_{n,\,3} \\ P_{n+1,\,3} &= A_3P_{n,\,1} + A_2P_{n,\,2} + A_1P_{n,\,3}. \end{aligned} \qquad (3)$$

Since the initial polynomials, $P_{0,\,i}$, are easily found, the $P_{n,\,i}$ are determined once the characteristic roots of the coefficient matrix

$$A = \begin{pmatrix} A_1 & A_3(1 + x) & A_2(1 + x) \\ A_2 & A_1 & A_3(1 + x) \\ A_3 & A_2 & A_1 \end{pmatrix} \qquad (4)$$

are known. This matrix resembles a circulant, and it is easy to verify that the characteristic roots are given by

$$\begin{aligned} \lambda_1 &= A_1 + A_2\sqrt[3]{1 + x} + A_3\sqrt[3]{(1 + x)^2} \\ \lambda_2 &= A_1 + \omega A_2\sqrt[3]{1 + x} + \omega^2A_3\sqrt[3]{(1 + x)^2} \\ \lambda_3 &= A_1 + \omega^2A_2\sqrt[3]{1 + x} + \omega A_3\sqrt[3]{(1 + x)^2}, \end{aligned} \qquad (5)$$

where $\omega$ is a primitive cubic root of unity.

These results generalize immediately to the $N$-dimensional case. The above formulas exhibit the known correspondence between the three-dimensional continued fraction and the decomposable cubic form

$$x^3 + y^3 + z^3 + 3xyz = (x + y + z)(x + \omega y + \omega^2 z)(x + \omega^2 y + \omega z) \qquad (6)$$

cf. Daus.[1]

The method sketched here for the three-dimensional case applies to those algebraic functions satisfying (3) of section 1.

---

[1] Daus, P. H., "Normal Ternary Continued Fraction Expansions for the Cube Roots of Integers," *Am. J. Math.*, **44**, 279 (1922).

[2] Perron, O., *Die Lehre von den Kettenbrüchen*, Chap. X, Leipzig, 1929.

[3] Whittaker, E. T., and Watson, G. N., *A Course of Modern Analysis*, chapters VIII, IX and X.

# ON THE NUMBER OF SOLUTIONS OF SOME TRINOMIAL EQUATIONS IN A FINITE FIELD

BY LOO-KENG HUA AND H. S. VANDIVER

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS, AND DEPARTMENT OF APPLIED MATHEMATICS, UNIVERSITY OF TEXAS

Communicated July 1, 1949

In the present paper we shall use the results in two previous papers[1] by the authors to obtain (Theorem I) the exact number, explicitly in terms of $p$, $n$, $k_1$ and $k_2$, of solutions of the equation, if $0 < a_i < p^n - 1$, $(p^n - 1, a_i) = k_i$, $i = 1, 2$;

$$c_1x_1^{a_1} + c_2x_2^{a_2} + c_3 = 0, \tag{1}$$

in $x_1$ and $x_2$, non-zero elements of a finite field $F(p^n)$ of order $p^n$, $p$ an odd prime with $c_1$, $c_2$, $c_3$ given elements of $F(p^n)$, $c_1c_2c_3 \neq 0$, in certain special cases. Secondly, we shall find limits (Theorem II) for the number of solutions of (1) which are better than those given in $C$ for the solutions of this particular equation, and which have the unusual property that they agree with the exact values found for the number of solutions in the special cases treated in Theorem I. We shall employ the notation used in H (defined in (9b) just below relation (14) of that paper).

Let $\alpha$ be a primitive $(ds_1s_2)$-th root of unity, $m = ds_1s_2$ where $ds_1 = k_1$, $ds_2 = k_2$ where $(k_1, k_2) = d$, whence $(s_1, s_2) = 1$. Let $\chi_{k_1}$ be a special $k_1$-character and $\chi_{k_2}$ a special $k_2$-character in $F(p^n)$ and write, if $K$ stands for $F(p^n)$,

$$\tau(\chi_{k_1}^{\mu_1}) = \sum_{a \in K} \alpha^{\mu_1 s_2 \operatorname{ind}(a)} \zeta^{tr(a)}$$

and

$$\tau(\chi_{k_2}^{\mu_2}) = \sum_{a \in K} \alpha^{\mu_2 s_1 \operatorname{ind}(a)} \zeta^{tr(a)},$$

where $tr(a)$ is the trace of $a$ in $F(p^n)$ with respect to $F(p)$, and $g^{\operatorname{ind}(a)}$ is defined by

$$g^{\operatorname{ind}(a)} = a$$

in $F(p^n)$, $g$ being a primitive root of $F(p^n)$ and we define ind (0) as zero. Hence, as is known, provided $\chi_{k_1}^{\mu_1}\chi_{k_2}^{\mu_2} \neq 1$,

$$\frac{\tau(\chi_{k_1}^{\mu_1})\tau(\chi_{k_2}^{\mu_2})}{\tau(\chi_{k_1}^{\mu_1}\,\chi_{k_2}^{\mu_2})} = \sum_{a} \alpha^{\mu_1 s_2 \operatorname{ind}(a) + \mu_2 s_1 \operatorname{ind}(1-a)}.$$

Also, setting $(-a_1)$ for $(a)$ we obtain

$$\begin{aligned}
&\sum_{a\,\epsilon\,K}\alpha^{\mu_1s_2\operatorname{ind}(a)+\mu_2s_1\operatorname{ind}(1-a)}\\
&=\sum_{a_1\,\epsilon\,K}\alpha^{\mu_1s_2\operatorname{ind}(-a_1)+\mu_2s_1\operatorname{ind}(1+a_1)}\\
&=\sum_{a_1\,\epsilon\,K}\alpha^{\mu_1s_2\operatorname{ind}(-1)}\cdot\alpha^{\mu_1s_2\operatorname{ind}(a_1)+\mu_2s_1\operatorname{ind}(1+a_1)}.
\end{aligned}$$

Let $n = 2n_1$, then $p^{2n_1} - 1 \equiv 0 \pmod 4$ and if $(\mu_1s_2(p^{2n_1} - 1))/2 = e$ so that $\alpha^{\mu_1s_2\operatorname{ind}(-1)} = (-1)^e = 1$,

$$\frac{\tau(\chi_{k_1}^{\mu_1})\tau(\chi_{k_2}^{\mu_2})}{\tau(\chi_{k_1}^{\mu_1}\chi_{k_2}^{\mu_2})} = \sum_{a\,\epsilon\,K}\alpha^{\mu_1s_2\operatorname{ind}(a)+\mu_2s_1\operatorname{ind}(a+1)} = \psi(\alpha) \tag{2}$$

Now Mitchell[2] showed that if $\mu_1s_2 \not\equiv 0 \pmod m$, $\mu_2s_1 \not\equiv 0 \pmod m$ and $\mu s_2 + \mu' s_1 \not\equiv 0 \pmod m$ then

$$\psi(\alpha) = (-1)^{r-1}p^{n_1}, \tag{3}$$

where $p$ belongs to the exponent $2t$, modulo $m$ and $n_1 = rt$ with $p^t \equiv -1 \pmod m$. We now employ (18) of H, which may be written as follows:

$$\begin{aligned}
N = p^n - 2 - \sum_{\mu_1=0}^{k_1-1}\chi_{k_1}^{\mu_1}\left(-\frac{c_3}{c_1}\right) - \sum_{\mu_2=0}^{k_2-1}\chi_{k_2}^{\mu_2}\left(-\frac{c_3}{c_2}\right) - \sum_{\nu=1}^{d-1}\chi_{k_1}^{s_1\nu}\left(-\frac{c_2}{c_1}\right)\\
+ \sum_{\mu_1,\,\mu_2}\chi_{k_1}^{\mu_1}\left(-\frac{c_3}{c_1}\right)\chi_{k_2}^{\mu_2}\left(-\frac{c_3}{c_2}\right)\frac{\tau(\chi_{k_1}^{\mu_1})\tau(\chi_{k_2}^{\mu_2})}{\tau(\chi_{k_1}^{\mu_1}\chi_{k_2}^{\mu_2})},
\end{aligned}\tag{4}$$

where in the last term $\chi_{k_1}^{\mu_1} \neq 1$, $\chi_{k_2}^{\mu_2} \neq 1$, $\chi_1^{\mu_1}\chi_2^{\mu_2} \neq 1$, and divide the discussion into seven cases in order to express $s$ in terms of integers only. Because of (3) and $\chi(\alpha) = -1$ for $\mu_1 = 0$, $\mu_2 \neq 0$, also for $\mu_1 \neq 0$, $\mu_2 = 0$, this is always possible.

The cases are:

I. $c_3 = c_2 = c_1$. II. $c_3 = c_1$, $c_2 \neq c_1$ with $\operatorname{ind} c_2 \equiv \operatorname{ind} c_1 \pmod d$. III. $c_3 = c_1$, $\operatorname{ind} c_2 \not\equiv \operatorname{ind} c_1 \pmod d$. IV. $c_3 = c_2$, $c_1 \neq c_2$, $\operatorname{ind} c_2 \equiv \operatorname{ind} c_1 \pmod d$. V. $c_3 = c_2$, $c_2 \not\equiv c_1 \pmod d$. VI. $c_3 \neq c_2$, $c_3 \neq c_1$, $\operatorname{ind} c_2 \equiv \operatorname{ind} c_1 \pmod d$. VII. $c_3 \neq c_1$, $c_3 \neq c_2$, $c_1 \not\equiv c_2 \pmod d$. Write

$$A = \sum_{\mu_1,\,\mu_2}\chi_{k_1}^{\mu_1}\left(-\frac{c_3}{c_1}\right)\chi_{k_2}^{\mu_2}\left(-\frac{c_3}{c_2}\right)\frac{\tau(\chi_{k_1}^{\mu_1})\tau(\chi_{k_2}^{\mu_2})}{\tau(\chi_{k_1}^{\mu_1}\chi_{k_2}^{\mu_2})} \tag{5}$$

where $\chi_{k_1}^{\mu_1} \neq 1$, $\chi_{k_2}^{\mu_2} \neq 1$, $\chi_{k_1}^{\mu_1}\chi_{k_2}^{\mu_2} \neq 1$.

In case I, we find

$$\sum_{\mu_1=0}^{k_1-1}\chi_{k_1}^{\mu_1}\left(-\frac{c_3}{c_1}\right) = k_1,$$

$$\sum_{\mu_2=0}^{k_2-1}\chi_{k_2}^{\mu_2}\left(-\frac{c_3}{c_2}\right) = k_2,$$

$$\sum_{v=1}^{d-1} \chi_{k_1}^{s_1 v}\left(-\frac{c_2}{c_1}\right) = d - 1.$$

To reduce $A$ in this case, we take all the terms in which $\chi_{k_1}^{\mu_1} \neq 1$, $\chi_{k_2}^{\mu_2} \neq 1$ and then subtract all the terms from the result in which $\chi_{k_1}^{\mu_1}\chi_{k_2}^{\mu_2} = 1$ with $\chi_{k_1}^{\mu_1} \neq 1$, $\chi_{k_2}^{\mu_2} \neq 1$ and employ (3). The terms of the first type give

$$(k_1 - 1)(k_2 - 1)p^{n_1}(-1)^{r-1}.$$

As in the proof of (18) of $s$ there are $(d - 1)$ terms of the second type just mentioned. Hence (4) reduces to

$$p^n + 1 - k_1 - k_2 - d + ((k_1 - 1)(k_2 - 1) - (d - 1))(-1)^{r-1}p^{n_1}. \tag{6}$$

For case II, we find the same value for the first summation in (4) but for the second, we have the value 0 since $c_3 \neq c_2$. In reducing $A$ for this case and separating the summation in the same way it was separated in case I, we have for the first part $(k_1 - 1)p^{n_1}(-1)^r$ and for the second part $(d - 1)p^{n_1}(-1)^r$. Hence, we have for case II the value, if $N_i$ denotes the number of solutions in case $i$, $i$ = I, II, ..., VII.

$$N_2 = p^n + 1 - d - k_1 + (k_1 + d - 2)p^{n_1}(-1)^r. \tag{7}$$

Similarly for cases III, IV, etc., we have

$$N_3 = p^n + 1 - k_1 + (k_1 - 2)p^{n_1}(-1)^r. \tag{8}$$

$$N_4 = p^n + 1 - d - k_2 + (k_2 + d - 2)p^{n_1}(-1)^r. \tag{9}$$

$$N_5 = p^n + 1 - k_2 + (k_2 - 2)p^{n_1}(-1)^r. \tag{10}$$

$$N_6 = p^n + 1 - d + p^{n_1}(-1)^r(d - 2). \tag{11}$$

$$N_7 = p^n + 1 + (-1)^{r-1}\, 2p^{n_1}. \tag{12}$$

THEOREM I. *If $N_i$ is the number of solutions of the equation*

$$c_1x_1^{a_1} + c_2x_2^{a_2} + c_3 = 0$$

*in $x_1$ and $x_2$, non-zero elements of a finite field $F(p^n)$ of order $p^n$, $p$ an odd prime, with $c_1$, $c_2$, $c_3$ given non-zero elements of $F(p^n)$ with $N_i$, $i = 1, 2, \ldots, 7$, the number of solutions in the corresponding seven cases numbered I, II, ..., VII, given just above the relation (5); $0 < a_i < p^n - 1$, $(a_i, p^n - 1) = k_i$, $i = 1, 2$; $m = ds_1 s_2$, $ds_1 = k_1$, $ds_2 = k_2$, $(k_1, k_2) = d$; $n = 2n_1$, $p$ belongs to the exponent $2t$ modulo $m$, $n_1 = rt$ with $p^t \equiv -1 \pmod{m}$, then the values of $N_i$ are given by the relations (6)–(12), inclusive.*

Mitchell[2] obtained these results for the special case when $k_1 = k_2$ by a more complicated method.

We now use (4) to find closer limits for $N$ in the special case when $s = 2$, than those found in C for equation (1) of that paper. Set

$$M = p^n \sum_{\mu_1=0}^{k_1-1} \chi_{k_1}^{\mu_1}\left(-\frac{c_3}{c_1}\right) - \sum_{\mu_2=0}^{k_2-1} \chi_{k_2}^{\mu_2}\left(-\frac{c_3}{c_2}\right) - \sum_{v=1}^{d-1} \chi_{k_1}^{v s_1}\left(-\frac{c_2}{c_1}\right).$$

Now every term on the right of (4) is real so that we may write $N \leqq M + |A|$ and $N \geqq M - |A|$ and using the known result

$$|\tau(\chi)| = p^{n/2},\ \chi \neq \chi_0, \tag{13}$$

we have, if

$$D = |((k_1 - 1)(k_2 - 1) - (d - 1))p^{n/2}|, \tag{13a}$$

$$N \leqq M + D, \tag{14}$$

$$N \geqq M - D. \tag{15}$$

To obtain limits in terms of rational integers only, divide the discussion into the seven cases I–VII defined in the proof of Theorem I. Then we can evaluate $M$ exactly as in the determination of $N$ in Theorem I in each of the seven cases.[4]

Write $N_i$; $i = 1, 2, \ldots, 7$ for the value of $N$ corresponding to each of the seven cases referred to in Theorem I, respectively. Then from (14) and (15) we have the

THEOREM II. *If*

$$c_1 x_1^{a_1} + c_2 x_2^{a_2} + c_3 = 0,$$

*where* $c_1, c_2, c_3$ *are given elements in a finite field* $F(p^n)$ *of order* $p^n$, $p$ *prime*, $c_1c_2c_3 \neq 0$, *then we have the following limits for* $N_i$, *the number of solutions in the x's, neither zero, of the above equation,*

$$M_i - D \leqq N_i \leqq M_i + D, \tag{16}$$

*where* $M_1 = p^n + 1 - k_1 - k_2 - d$; $M_2 = p^n + 1 - d - k_1$; $M_3 = p^n + 1 - k_1$; $M_4 = p^n + 1 - d - k_2$; $M_5 = p^n + 1 - k_2$; $M_6 = p^n + 1 - d$; $M_7 = p^n + 1$; $k_1 = (p^n - 1, a_1)$, $k_2 = (p^n - 1, a_2)$, $d = (k_1, k_2)$ *and* $D$ *is defined in* (13a), $M_i$; $i = 1, 2, \ldots, 7$, *is the value of* $M$ *corresponding to each of the seven cases, I, II, ..., VII, of Theorem I, respectively.*

For the case when $n = 1$ and $k_1 = k_2$ is an odd prime, Hurwitz[3] by a different method proved the above theorem using rational integers only. His argument is much longer than ours. For the particular cases treated in Theorem I the *exact value for* $N$ *agrees with the limits found in Theorem II, when* $c_1 = c_2 = c_3$, *with the superior limit when* $r$ *is odd and the inferior limit when* $r$ *is even*. For the case $n = 1$, $k_1 = k_2$ an odd prime, this fact was noted by Mitchell,[2] who also stated that Hurwitz's method could be extended so as to obtain limits for the case when $n$ is general, which had a similar property.

[1] These PROCEEDINGS, **35**, 94–99 (1949), this will be referred to as C; **35**, 451–457 (1949), this will be referred to as H.

[2] *Ann. Math.*, **II**, 18, 120 (1917).

[3] *Crelle*, **136**, 272–292 (1909).

[4] As a check on the accuracy of our work, the present problem was discussed in a mathematical seminar, using a different method of proof. During the resulting discussion, Mr. Olin B. Faircloth made a suggestion, which, when used in connection with the method employed in this paper yielded a better superior limit for $N$, than that which we originally had.

---

# ON THE NATURE OF THE SOLUTIONS OF CERTAIN EQUATIONS IN A FINITE FIELD

BY L. K. HUA AND H. S. VANDIVER

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS, AND DEPARTMENT OF APPLIED MATHEMATICS, UNIVERSITY OF TEXAS

Communicated May 23, 1949

In two recent papers[1] the authors considered the equation

$$c_1x_1^{a_1} + c_2x_2^{a_2} + \ldots + c_sx_s^{a_s} + c_{s+1} = 0 \qquad (1)$$

in the $x$'s, where the $a$'s are integers such that $0 < a < p^n - 1$; $s \geqq 2$ for $c_{s+1} \neq 0$, and $s > 2$ for $c_{s+1} = 0$ the $c$'s being given elements of a finite field of order $p^n$, $p$ an odd prime, which will be designated by $F(p^n)$; and

$$c_1c_2 \ldots c_sx_1x_2 \ldots x_s \neq 0 \qquad (2)$$

in $F(p^n)$. Limits were found for the number of solutions of (1) and it was proved that for $p^n$ sufficiently large (1) always had solutions under the restriction (2). In the statement above concerning (1) we made the provision that $s \geqq 2$ for $c_{s+1} \neq 0$. If we consider the case where $s = 1$, (1) reduces to the equation

$$c_1x_1^{a_1} + c_2 = 0, \qquad (3)$$

with $c_1c_2 \neq 0$. Since any finite field can be represented by means of the residue classes with respect to some prime ideal in an algebraic field, then examination of the solutions of (3) is included in the theory of the congruence

$$x^n \equiv \alpha \pmod{\mathfrak{p}}, \qquad (4)$$

where $\alpha$ is an integer in an algebraic field $K$, $\mathfrak{p}$ is a prime ideal in that field and $\alpha \not\equiv 0 \pmod{\mathfrak{p}}$. The study of the last relation led to the theory of the class field and the laws of reciprocity for $n$th powers.[2]

*We wish to point out the sharp distinction between this problem and the problem of finding the solutions or the number of solutions of the equation (1) and when* $s \geqq 2$ for $c_{s+1} \neq 0$.

In addition to this, we shall also obtain an expression (Theorem I) for the number of solutions of (1) when $c_{s+1} \neq 0$ and $s = 2$, which leads to the proof of Theorem I and which we shall also have occasion to make use of in another paper. Also we obtain Theorem II, which gives the exact number of solutions of (1) when the $k$'s are prime each to each, $c_{s+1} = 0$ and $(a_i, p^n - 1) = k_i$; $i = 1, 2, \ldots, s$.

We consider equation (4) and confine ourselves to the case where $K$ is an algebraic field which contains a primitive $l$th root of unity with $l$ prime. Then

$$x^l - \alpha \equiv 0 \pmod{\mathfrak{p}} \tag{5}$$

if and only if the $l$th power character of $\alpha$, modulo $\mathfrak{p}$ is unity, or

$$\left(\frac{\alpha}{\mathfrak{p}}\right)_l = 1.$$

However, by a known theorem[3]

$$\left(\frac{\alpha}{\mathfrak{p}}\right)_l \neq 1$$

for an infinity of distinct prime ideals $\mathfrak{p}$ provided $\alpha$ does not equal an $l$th power of an integer in $K$. Now the last statement is contrary to what we have for $s > 1$ in (1) with $c_{s+1} \neq 0$. For, as we have shown in $E$ (p. 262) the congruence

$$\alpha_1 x_1^{a_1} + \alpha_2 x_2^{a_2} + \ldots + \alpha_s x_s^{a_s} + \alpha_{s+1} \equiv 0 \pmod{\mathfrak{p}}, \tag{6}$$

with

$$\alpha_1 \alpha_2 \ldots \alpha_s x_1 x_2 \ldots x_s \not\equiv 0 \pmod{\mathfrak{p}}$$

for $s \geqq 2$ with $c_{s+1} \not\equiv 0 \pmod{\mathfrak{p}}$ always has solutions if the norm of $\mathfrak{p}$ exceeds a certain limit, in fact (6) always has at least $k$ solutions in the $x$'s for $k$ any positive integer if the norm of $p$ exceeds a certain limit. In particular when we consider the case where there are no solutions it is known[4] that $x^7 + y^7 + 1 \equiv 0 \pmod{\rho}$ has no solutions in integers prime to $p$ for just four primes $p$, namely 29, 71, 113, 491, but has solutions for all the other primes $p$. As for $x^5 + y^5 + 1 \equiv 0 \pmod{p}$ it has no solutions for $p = 11, 41, 71, 101$. Also $x^3 + y^3 + 1 \equiv 0 \pmod{p}$ has no solutions only for $p = 7$ and 13. It is also known[5] that for $m < 10l$, and $p = 1 + ml$ and $h(l)$ is the number of primes $p$ for which

$$x^l + y^l + 1 \equiv 0 \pmod{p} \tag{7}$$

has no solutions for $xy \not\equiv 0 \pmod{p}$ then $h(11) = 3$, $h(13) = 4$, $h(17) = 4$,

$h(19) = 6$. In view of these facts it may happen that $h(l)$ increases as $l$ increases. Note also from the cases 3, 5 and 7 that there is a range of values for $p$ in which the corresponding congruences may or may not have solutions.

From the above considerations we may say that associated with any form of the type given by the left-hand member of (4) there is an integer $e(k)$ such that there are exactly $e(k)$ values of $\mathfrak{p}$ such that (4) has $k$ solutions. Even in the case where $k$ is zero, the class of forms satisfying the condition may be rather general. For example,[6] it was shown that if $c$ is a prime and $m$ is an integer such that $p = 1 + mc$ with $p$ prime then

$$a_1x_1^m + a_2x_2^m + \ldots + a_sx_s^m \equiv 0 \pmod{p}$$

has only the solutions $x_1 \equiv x_2 \equiv \ldots \equiv x_s \equiv 0 \pmod{p}$ provided $s \leqq c - 2$, the sum of no $n$ of the $a$'s is $\equiv 0 \pmod{p}$ $0 < n \leqq s$, and

$$(|a_1| + |a_2| + \ldots + |a_s|)^{\phi(c)} < p,$$

where $\phi(c)$ is the indicator of $c$. Select $a_1 = a_2 = \ldots = a_{s-1} = 1$ and $a_s = -s$, and also an $m$ and $p$ with $p = 1 + mc$ with $c$ fixed and satisfying the other conditions just mentioned then it is evident that

$$x_1^m + x_2^m + \ldots + x_{s-1}^m - sx_s^m \equiv 0 \pmod{p}$$

has no primitive solutions. (Incidentally, in view of Theorem III of the article just referred to, the equation

$$x_1^m + x_2^m + \ldots + x_{s-1}^m - sx_s^m = 0$$

has no solutions in rational integers unless each $x$ is zero, if there is a prime $p$ with $p = 1 + mc$ and $(2s - 1)^{\varphi(c)} < p$.)

We now recur to the subject of the number of solutions of equation (1), treated in E and C. The case $s = 2$ with $c_3 \neq 0$, has received considerable attention in the literature.[7] Let us examine

$$c_1x_1^{a_1} + c_2x_2^{a_2} + c_3 = 0, \tag{8}$$

with

$$x_1x_2c_1c_2c_3 \neq 0. \tag{8a}$$

Theorem I of C, we see will give, if $N$ is the number of solutions of (8) in $x_1$ and $x_2$ under the conditions (8a),

$$N = \frac{(p^n - 1)^2}{p^n} + \frac{T}{p^n}, \tag{9}$$

where we may write $T$ in the form, if $k_i = (a_i, p - 1)$, $\mu_1$ ranges over 0, 1, ..., $k_1 - 1$; $\mu_2$ over 0, 1, ..., $k_2 - 1$

$$\sum_{\substack{a \in K \\ a \neq 0}} \sum_{\mu_1, \mu_2} \tau(\chi_{k_1}^{\mu_1})\tau(\chi_{k_2}^{\mu_2})\chi_{k_1}^{\mu_1}(c_1^{-1})\chi_{k_2}^{\mu_2}(c_2^{-1})\chi_{k_1}^{\mu_1}(a^{-1})\chi_{k_2}^{\mu_2}(a^{-1})\zeta^{tr(ac_3)}, \tag{9a}$$

where $\chi_{k_1}$ is a special character defined by $k_1$ or as we shall say, a $k_1$-character, and similarly $\chi_{k_2}$ is a special character defined by $k_2$ and where, if we write $K$ for $F(p^n)$,

$$\tau(\chi, \zeta) = \tau(\chi) = \sum_{b \epsilon K} \chi(b)\zeta^{tr(b)}; \ \zeta = e^{2i\pi/p}, \tag{9b}$$

with $\chi(0) = 0$. This is obtained from the fact that if $\alpha$ is a primitive $(p^n - 1)$th root of unity then $\alpha^{\mu \text{ ind } a}$, where $a$ ranges over all elements of $F(p^n)$, $\neq 0$, gives a character as defined in C where ind $a$ is given by

$$g^{\text{ind } a} = a,$$

and where $g$ is a primitive root in $F(p^n)$, so we obtain all characters by taking $\mu = 0, 1, \ldots, p^n - 2$. We have from C the relations, if $\chi_0$ is the principal character,

$$\left.\begin{array}{lll} |\tau(\chi)| & = & p^{n/2}; \ \chi \neq \chi_0 \\ \tau(\chi_0) & = & -1 \end{array}\right\}. \tag{10}$$

To reduce (9a) we divide the discussion into four cases.

First Case. $\chi_{k_1}^{\mu_1} = 1, \chi_{k_2}^{\mu_2} = 1$. In this event $\chi_{k_1}^{\mu_1}(a^{-1}) = \chi_{k_2}^{\mu_2}(a^{-1}) = 1$, $\chi_{k_1}^{\mu_1}(c_1^{-1}) = \chi_{k_2}^{\mu_2}(c_2^{-1}) = 1$, which reduces (9a), using (10), to

$$\sum_a \zeta^{tr(ac_3)} = -1. \tag{11}$$

Second Case. $\chi_{k_1}^{\mu_1} = 1; \ \chi_{k_2}^{\mu_2} \neq 1$. This gives from (9a) and (10),

$$-\sum_a \tau(\chi_{k_2}^{\mu_2})\chi_{k_2}^{\mu_2}(c_2^{-1}a^{-1})\zeta^{tr(ac_3)} = \sum_a - \tau(\chi_{k_2})\chi_{k_2}(c_2^{-1})(\chi_{k_2}(a))^{-1}\zeta^{tr(ac_3)},$$

and setting $ac_3 = -a_1$, using $c_3 \neq 0$, this reduces to

$$-\sum_{a_1} \tau(\chi_{k_2}^{\mu_2})\chi_{k_2}^{\mu_2}(-c_2^{-1}c_3)(\chi_{k_2}^{\mu_2}(a_1))^{-1}\zeta^{-tr\ a_1} = -\sum \tau(\chi_{k_2})\overline{\tau(\chi_{k_2})}\chi_{k_2}\left(-\frac{c_3}{c_2}\right),$$

and using the known relation

$$\tau(\chi_{k_2}^{\mu_2})\overline{\tau(\chi_{k_2}^{\mu_2})} = p^n, \tag{11a}$$

this is seen to equal

$$-p^n \sum_{\mu_2} \chi_{k_2}^{\mu_2}\left(-\frac{c_3}{c_2}\right). \tag{12}$$

Also, if $\chi_{k_1}^{\mu_1} \neq 1$ and $\chi_{k_2}^{\mu_2} = 1$ we find in a similar way that the result is

$$-p^n \sum_{\mu_1} \chi_{k_1}^{\mu_1}\left(-\frac{c_3}{c_1}\right). \tag{13}$$

Third Case. $\chi_{k_1}^{\mu_1}\chi_{k_2}^{\mu_2} = 1$ with $\chi_{k_1}^{\mu_1} \neq 1, \chi_{k_2}^{\mu_2} \neq 1$, (9a) reduces to

$$\sum_{a}\sum_{\mu_1}\chi_{k_1}^{\mu_1}(c_1^{-1})\chi_{k_1}^{-\mu_1}(c_2^{-1})\tau(\chi_{k_1}^{\mu_1})\tau(\chi_{k_1}^{-\mu_1}, \zeta)\zeta^{tr(ac_3)}.$$

Now

$$\begin{aligned}\tau(\chi_{k_1}^{-\mu_1}, \zeta) &= \tau(\chi_{k_1}^{-\mu_1}, \zeta^{-1})\chi_{k_1}(-1)\\ &= \overline{\tau(\chi_{k_1}^{\mu_1})}\chi_{k_1}^{\mu_1}(-1).\end{aligned}$$

Hence for the admissible values of $\mu_1$, (9a) reduces to

$$-\sum_{\mu_1}\chi_{k_1}^{\mu_1}\left(-\frac{c_2}{c_1}\right)p^n. \tag{14}$$

Now we must determine the possible values of $\mu_1$ from $\chi_{k_1}^{\mu_1}\chi_{k_2}^{\mu_2} = 1$. Let $\alpha$ be a primitive $(ds_1s_2)$th root of unity where $ds_1 = k_1$, $ds_2 = k_2$ and where $(k_1, k_2) = d$, whence $(s_1, s_2) = 1$. Then $\chi_{k_1}^{\mu_1}\chi_{k_2}^{\mu_2} = 1$ gives integers $\mu_1$ and $\mu_2$ such that $\alpha^{s_1\mu_1 + s_2\mu_2} = 1$ or $s_2\mu_1 + s_1\mu_2 \equiv 0 \pmod{(ds_1s_2)}$ so that $\mu_1 = v_1s_1$, $\mu_2 = v_2s_2$. Hence the possible values of $\mu_1$ are $s_1v_1$; $v_1 = 1, 2, \ldots, d - 1$, as the case $v_1 = 0$ was excepted. Hence (9a) reduces to, for this case,

$$-\sum_{v_1 = 1}^{d-1}\chi_{k_1}^{s_1v_1}\left(-\frac{c_2}{c_1}\right)p^n. \tag{15}$$

Fourth Case. Here $\chi_{k_1}^{\mu_1} \neq 1$, $\chi_{k_2}^{\mu_2} \neq 1$, $\chi_{k_1}^{\mu_1}\chi_{k_2}^{\mu_2} \neq 1$. Consider such terms in (9a). They may be written in the form

$$\sum_{\mu_1, \mu_2}\sum_{a \in K}\tau(\chi_{k_1}^{\mu_1})\tau(\chi_{k_2}^{\mu_2})\chi_{k_1}^{\mu_1}\chi_{k_2}^{\mu_2}(a^{-1})\zeta^{tr(ac_3)}\cdot\chi_{k_1}^{\mu_1}(c_1^{-1})\chi_{k_2}^{\mu_2}(c_2^{-1}). \tag{16}$$

Write $a_1 = ac_3$, then $a^{-1} = a_1^{-1}c_3$ and using $c_3 \neq 0$,

$$\begin{aligned}\sum_{\substack{a_1 \in K\\ a_1 \neq 0}}\chi_{k_1}^{\mu_1}\chi_{k_2}^{\mu_2}(a^{-1}) &= \sum_{a_1 \in K}\chi_{k_1}^{\mu_1}\chi_{k_2}^{\mu_2}(c_3)(\chi_{k_1}^{\mu_1}\chi_{k_2}^{\mu_2}(a))^{-1}\zeta^{tr(a_1)}\\ &= \sum_{a_1 \in K}\chi_{k_1}^{\mu_1}\chi_{k_2}^{\mu_2}(c_3)(\chi_{k_1}^{\mu_1}\chi_{k_2}^{\mu_2}(a_1))^{-1}\chi_{k_1}^{\mu_1}\chi_{k_2}^{\mu_2}(-1)\zeta^{-tr(a_1)}\end{aligned}$$

Hence using (11a), we find from (9a) that the terms in case 4 in $T$ may be written as

$$p^n\sum_{\mu_1, \mu_2}\chi_{k_1}^{\mu_1}\left(-\frac{c_3}{c_1}\right)\chi_{k_2}^{\mu_2}\left(-\frac{c_3}{c_2}\right)\frac{\tau(\chi_{k_1}^{\mu_1})(\chi_{k_2}^{\mu_2})}{\tau(\chi_{k_1}^{\mu_1}\chi_{k_2}^{\mu_2})}, \tag{17}$$

$\chi_{k_1}^{\mu_1} \neq 1$, $\chi_{k_2}^{\mu_2} \neq 1$, $\chi_{k_1}^{\mu_1}\chi_{k_2}^{\mu_2} \neq 1$. Using (9), (9a), (11), (12), (13), (14), (15) and (17), we find

$$\begin{aligned}N = p^n &- \sum_{\mu_1 = 0}^{k_1 - 1}\chi_{k_1}^{\mu_1}\left(-\frac{c_3}{c_1}\right) - \sum_{\mu_2 = 0}^{k_2 - 1}\chi_{k_2}^{\mu_2}\left(-\frac{c_3}{c_2}\right) - \sum_{v = 1}^{d - 1}\chi_{k_1}^{s_1v}\left(-\frac{c_2}{c_1}\right)\\ &+ \sum_{\mu_1, \mu_2}{}'\chi_{k_1}^{\mu_1}\left(-\frac{c_3}{c_1}\right)\chi_{k_2}^{\mu_2}\left(-\frac{c_3}{c_2}\right)\frac{\tau(\chi_{k_1}^{\mu_1})\tau(\chi_{k_2}^{\mu_2})}{\tau(\chi_{k_1}^{\mu_1}\chi_{k_2}^{\mu_2})}\end{aligned} \tag{18}$$

where, in the last term, $\chi_{k_1}^{\mu_1} \neq 1$, $\chi_{k_2}^{\mu_2} \neq 1$, $\chi_{k_1}^{\mu_1}\chi_{k_2}^{\mu_2} \neq 1$.

Whence we have the (cf. Davenport-Hasse[8])

THEOREM I. *If $N$ is the number of solutions in $x_1$ and $x_2$ of*

$$c_1x_1^{a_1} + c_2x_2^{a_2} + c_3 = 0,$$

*with $c_1$, $c_2$, $c_3$, $x_1$, $x_2$ in a finite field $F(p^n)$ of order $p^n$ with $p$ an odd prime, $c_1c_2c_3x_1x_2 \neq 0$ in $F(p^n)$; $(a_1, p^n - 1) = k_1$, $0 < k_1 < p^n - 1$, $(a_2, p^n - 1) = k_2$, $0 < k_2 < p^n - 1$, $\chi_{k_1}$ is a special $k_1$-character, $\chi_{k_2}$ is a special $k_2$-character, $d = (k_1, k_2)$, $ds_1 = k_1$, $\tau(\chi)$ is defined as in (9b), then the relation (18) holds.*

We shall now obtain $N$ explicitly in terms of $p$, $n$ and $s$, for special values of $k_1$ when $c_{s+1} = 0$. By the remark immediately following (9b), we may write

$$\chi_{k_i}^{\mu_i}(a^{-1}) = \alpha_i^{\mu_i \operatorname{ind} (a^{-1})} \tag{19}$$

for any $i$ in the set $0, 1, \ldots, k_i - 1$, and where $\alpha_i$ is a primitive $k_i$th root of unity. Set in Theorem I of $C$, $c_{s+1} = 0$ then for this case we have

$$T = \sum_{\mu_i = 0}^{k_i - 1} \sum_{\substack{a \in K \\ a \neq 0}} (\alpha_1^{\mu_1} \alpha_2^{\mu_2} \ldots \alpha_s^{\mu_s})^{\operatorname{ind} (a^{-1})} \cdot \prod_{i = 1}^{s} \alpha_i^{\mu_i \operatorname{ind} (c_i^{-1})} \tau(\alpha_i^{\mu_i}). \tag{20}$$

Assume that the $k$'s are prime each to each and carry out the summation with respect to $a$ we obtain

THEOREM II. *If $N_s$ is the number of solutions of (1) in $x$'s under the conditions (2) and the other conditions stated in connection with (1), and also $k_1, k_2, \ldots, k_s$ are prime each to each with $c_{s+1} = 0$, then*

$$N_s = \frac{p^n - 1}{p^n} ((p^n - 1)^{s-1} + (-1)^s). \tag{21}$$

This result may be proved in a simpler way as we shall now show.

Consider the equation

$$c_1x_1^{k_1} + \ldots + c_sx_s^{k_s} = 0, \quad (k_i, k_j) = 1, \quad i \neq j, \tag{22}$$

and $A = k_1k_2 \ldots k_s$ dividing $p^n - 1$. Since $(k_1, k_2, \ldots, k_s) = 1$, we have integers $\lambda$ and $\mu$ such that

$$\lambda k_1 + \mu k_2 \ldots k_s = 1, \qquad (\lambda, p^n - 1) = 1.$$

Putting

$$x_1 = y_1^{\lambda}, \quad x_2 = y_1^{-\mu A/k_1k_2}y_2, \quad x_3 = y_1^{-\mu A/k_1k_3}y_3, \ldots$$

Then we have

$$c_1y_1^{\lambda k_1} + y_1^{-\mu A/k_1}(c_2y_2^{k_2} + \ldots + c_sy_s^{k_s}) = 0$$

i.e.,

$$c_1y_1 + c_2y_2^{k_2} + \ldots + c_sy_s^{k_s} = 0$$

Since, for given elements of $F(p^n)$, $y_2, \ldots, y_s$ such that

$$c_2y_2^{k_2} + \ldots + c_sy_s^{k_s} \neq 0$$

we have a unique $y_1$, therefore, we have

$$N_s = (p^n - 1)^{s-1} - N_{s-1}.$$

Consequently

$$N_s = (p^n - 1)^{s-1} - (p^n - 1)^{s-2} + \ldots + (-1)^s(p^n - 1)$$
$$= \frac{p^n - 1}{p^n}\,((p^n - 1)^{s-1} + (-1)^s).$$

We also have, since the residue classes with respect to a rational prime modulus $p$ form a finite field of order $p$, the

COROLLARY I. *If $N$ is the number of incongruent solutions in the y's of the congruence, with $p$ an odd prime,*

$$d_1y_1^{a_1} + d_2y^{a_2} + \ldots + d_sy_s^{a_s} \equiv 0 \pmod{p}$$

*where $d_1d_2 \ldots d_s \not\equiv 0 \pmod{p}$, $y_1y_2 \ldots y_s \not\equiv 0 \pmod{p}$; $k_1, k_2, \ldots, k_s$, prime each to each with $(a_i, p - 1) = k_i$; $0 < k_i < p - 1$, $i = 1, 2, \ldots, s$ then*

$$N = \frac{p - 1}{p}\,((p - 1)^{s-1} + (-1)^s).$$

It may also be noted that the procedure used in the second proof of theorem II gives us a method for determining the solutions of (22).

---

[1] These PROCEEDINGS, **34**, 258–263 (1948), this will be referred to as E; **35**, 94–99 (1949), this will be referred to as C.

[2] Hasse, H., *Ber. Deut. Math. Verein.*, **35**, 1–55 (1926); **36**, 233–311 (1927); *Erganzungsband*, **VI** (1930).

[3] Takagi, T., *J. Coll. Sci., Imp. Univ. Tokyo*, **42**, 16 (1920).

[4] Dickson, L. E., *Messenger Math.*, **38**, 14–33 (1908).

[5] These results are due to N. G. W. H. Beeger, unpublished.

[6] Vandiver, H. S., these PROCEEDINGS, **32**, 101–106 (1946).

[7] Cf. footnote 1 of E.

[8] The proof of this theorem we gave as a direct application of our general theorem I of C to the special case $s = 2$. Another proof may be obtained by using the result of Davenport and Hasse, *Jour. für Math.* (Crelle) 172, p. 174 relation (6.5) which gives the number of solutions of $ax^m + by^n = c$, $abc \neq 0$ in $F(p^n)$, where $x$ and $y$ are not restricted, as they were in our work, to non-zero values in $F(p^n)$. To do this we consider the possible solutions involving zero and subtract the number of them from the right-hand member of the relation (6.5) just referred to. We note that the number of solutions of $ax^m = c$ and $by^n = c$ are, respectively,

$$\sum_{\mu=0}^{m-1} \bar{\chi}^{\mu}\left(\frac{c}{a}\right); \sum_{\nu=0}^{n-1} \psi\left(\frac{c}{b}\right)$$

which when taken in connection with (6.5) yields a formula equivalent to our relation (18).

# SUPERCONDUCTIVITY AND THE BOHR MAGNETON

By E. U. Condon

National Bureau of Standards, Washington, D. C.

Communicated June 15, 1949

Each superconducting metal is characterized by a function $H_c(T)$ where $H_c$ is a magnetic field in gauss and $T$ is absolute temperature, which determines the boundary between those states in which the material shows superconducting properties and the region where the properties are normal. That is, for each $T$, the magnetic field $H$ must be less than $H_c(T)$ for the metal to be in the superconducting state.

This functional relation plays an important rôle in the thermodynamics of superconductivity. If $G_n$ is the Gibbs free energy of unit volume of the metal in its normal state and $G_s$ is that in the superconducting state, but in zero magnetic field, then the Gibbs free energy of the superconductor in a magnetic field $H$ is $G_s + H^2/8\pi$. At the field $H = H_c$ the normal and superconducting phases are in equilibrium so $G_n = G_s + H_c^2/8\pi$. This relation is the starting point of the Keesom-Rutgers-Gorter analysis of thermodynamic properties of superconductors, which has been so extraordinarily fruitful in the last decade.

The general form of $H_c(T)$ is usually that given by the formula

$$H_c(T) = a(T_0^2 - T^2)$$

which forms a basis for extrapolation of the observational data to $T = 0$. There is a nice graph of the data in figure 23 of Schoenberg's book on Superconductivity.[1] Aside from the question of the exact validity of this particular formula, the boundary curve has two particularly interesting parameters, namely, the critical field at absolute zero, $H_c(0)$, and the temperature $T_0$ for which $H_c$ is zero.

If a long cylinder of superconducting material is magnetized by a field parallel to its axis it is known to behave as a perfect diamagnetic so that $B = 0$ inside it except for a skin layer of the order of $10^{-6}$ cm. thickness in which $B$ drops from the external $H$ down to zero in virtue of the building up of a $4\pi I$ oppositely directed to $H$ and equal to it in magnitude when one gets below the skin depth. From this point of view, $I_c = H_c/4\pi$ is the maximum amount of magnetic moment per unit volume which can be set up in the superconductor without destruction of the superconducting state.

When one focusses attention on $I_c = H_c/4\pi$ as a limiting magnetic moment per unit volume it is natural to inquire whether empirically there is any relation between this number and the product $N\beta$ where $N$ is the number of atoms per unit volume of the metal and $\beta$ is the Bohr magneton.

No reason at all is offered for doing this except that these are of like dimensions and are both related to fundamental magnetic properties.

The results are shown in the table, the values of $H_c(0)$ having been read by rough extrapolation of the curves in Schoenberg's book and the values of $N$ being calculated from standard density and atomic weight data.

TABLE 1

| ELEMENT | $I_0 = N\beta$ | $I_c(0) = H_c(0)/4\pi$ | $K^{-1} = I_0/I_c$ |
|---|---|---|---|
| Pb | 306 | 63 | 4.8 |
| Ta | 513 | 92 | 5.6 |
| Sn | 271 | 26 | 10.4 |
| Hg | 376 | 35.8 | 10.5 |
| In | 356 | 23.1 | 15.4 |
| Tl | 324 | 15.9 | 20.4 |
| Cb | 506 | 207 | 2.5 |
| Zr | 393 | 17.9 | 21.9 |
| Cd* | 430 | 4 | 117 |
| Al* | 559 | 14 | 40 |

* The data for the last two, Cd and Al, are so uncertain that it is doubtful whether they should be included.

It seems remarkable that in all cases the quantities, $I_0$ and $I_c$, are of the same general order of magnitude. It is also striking that the values of $I_0/I_c$ are roughly close to being an integral multiple of 5 (except columbium which requires a half quantum number!) so that the data here presented points roughly to a possible empirical relation,

$$I_c(0) = KN\beta, \tag{1}$$

in which, approximately, $K = {}^1/_{5n}$, where $n$ is an integer (except that $n = {}^1/_2$ for columbium) characteristic of the metal. Whether this integer rule will stand up to a more critical study of the data remains to be seen.

It is interesting to compare the relation of $I_c$ to $N\beta$ with the relation between mass transport of electrons by the superconducting electrons and Planck's angular momentum unit which was discovered by F. London.[2] London remarks that the electromagnetic measure of current per unit length of the cylinder wall, integrated through the skin depth is equal to $H/4\pi$ abamp./cm. where $H$ is in gauss. To find the maximum mass transport in gram/cm. sec., $H_c/4\pi$ must therefore be multiplied by $mc/e$. London next remarks that this has the same dimensions as $n_s\hbar$ where $\hbar$ is the Planck unit of angular momentum and $n_s$ is the number of superconducting electrons in unit volume, and then proceeds to present some evidence that these are of the same order of magnitude.

London's relation is thus

$$mcH/4\pi e = K'n_s\hbar$$

where $K'$ is an empirical coefficient which he finds to be of the order of unity on the basis of some scanty data for lead and tin.

In view of the fact that the Bohr magneton, $\beta$, is related to the fundamental constants by $\beta = e\hbar/2mc$, London's relation can be written

$$I_c = H_c/4\pi = 2K'n_e\beta \qquad (2)$$

so that it is very closely similar to the relation given in equation (1). The two relations can be put together in such a way as to give an expression for the atomic concentration of superconducting electrons as follows:

$$n_e/N = K/2K'. \qquad (3)$$

Of course it is not really necessary to go through these relations to get a value of $n_e/N$. Basically, one finds $n_e$, in terms of London's theory, by making some kind of experimental determination of the skin-depth parameter, $\lambda$, from which $n_e$ is calculated by the relation, $n_e = (mc^2/e^2)/4\pi\lambda^2$. The real difficulty lies in experimental measurement of $\lambda$. For this the available data are sparse and uncertain.

In London's paper[2] he only presents rough data for mercury and tin which lead him to say that $K' \sim 1$ for $Hg$ and $K' \sim {}^1/_6$ for Sn. Taking these values in combination with the values of $K$ given in table 1, one gets

$$n_e = {}^1/_{20} \text{ for } Hg \text{ and } n_e/N = {}^1/_3 \text{ for Sn.}$$

These are rather larger values than are usually supposed to be applicable, which may indicate that the experimental values for $\lambda$ used by London are rather too small.

These considerations strongly point to the desirability of obtaining more accurate measurements of $\lambda$ and of $H_c(0)$ for as many superconductors as possible.

[1] Schoenberg, D., *Superconductivity*, Cambridge University Press, 1938.

[2] London, F., *Rev. Mod. Phys.*, 17, 310 (1945).

---

# SEPARATION OF THE VARIABLES OF THE TWO-PARTICLE WAVE EQUATION

BY LUTHER PFAHLER EISENHART

FINE HALL, PRINCETON, N. J.

Communicated July 1, 1949

The wave equation of two particles at points $P(x, y, z)$ and $\bar{P}(\bar{x}, \bar{y}, \bar{z})$, where $x$, $y$, $z$ and $\bar{x}$, $\bar{y}$, $\bar{z}$ are rectangular-cartesian coördinates in euclidean 3-space is

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} + \frac{\partial^2}{\partial \bar{x}^2} + \frac{\partial^2}{\partial \bar{y}^2} + \frac{\partial^2}{\partial \bar{z}^2}\right)\psi + (E - V)\psi = 0,$$

where $V$ is the potential. The distances $r_1$ and $r_2$ of the particles from the origin and their mutual distance from one another $r_3$ are given by

$$r_1^2 = x^2 + y^2 + z^2, \qquad r_2^2 = \bar{x}^2 + \bar{y}^2 + \bar{z}^2, \qquad r_3^2 = \sum(x - \bar{x})^2. \tag{1}$$

Hylleraas[1] showed that for a solution $\psi(r_1, r_2, r_3)$ the above equation reduces to

$$\frac{\partial^2\psi}{\partial r_1^2} + \frac{\partial^2\psi}{\partial r_2^1} + 2\frac{\partial^2\psi}{\partial r_3^2} + \frac{r_1^2 - r_2^2 + r_3^2}{r_1 r_3}\frac{\partial^2\psi}{\partial r_1 \partial r_3} + \frac{r_2^2 - r_1^2 + r_3^2}{r_2 r_3}\frac{\partial^2\psi}{\partial r_2 \partial r_3} + \frac{2}{r_1}\frac{\partial\psi}{\partial r_1} + \frac{2}{r_2}\frac{\partial\psi}{\partial r_2} + \frac{4}{r_3}\frac{\partial\psi}{\partial r_3} + (E - V)\psi = 0. \tag{2}$$

If we put

$$x_1 = R_1, \qquad x_2 = R_2, \qquad x_3 = f(r_1, r_2, r_3), \tag{3}$$

where $R_1$ and $R_2$ are functions of $r_1$ and $r_2$ respectively, in the coördinates $x_i$ the above equation is

$$\begin{aligned} R_1'^{\,2}\frac{\partial^2\psi}{\partial x_1^2} + R_2'^{\,2}\frac{\partial^2\psi}{\partial x_2^2} + \Bigg[\left(\frac{\partial f}{\partial r_1}\right)^2 + \left(\frac{\partial f}{\partial r_2}\right)^2 + 2\left(\frac{\partial f}{\partial r_3}\right)^2 + \\ \frac{r_1^2 - r_2^2 + r_3^2}{r_1 r_3}\frac{\partial f}{\partial r_1}\frac{\partial f}{\partial r_3} + \frac{r_2^2 - r_1^2 + r_3^2}{r_2 r_3}\frac{\partial f}{\partial r_2}\frac{\partial f}{\partial r_3}\Bigg]\frac{\partial^2\psi}{\partial x_3^2} + \\ 2R_1'\left(\frac{\partial f}{\partial r_1} + \frac{r_1^2 - r_2^2 + r_3^2}{2r_1 r_3}\frac{\partial f}{\partial r_3}\right)\frac{\partial^2\psi}{\partial x_1 \partial x_3} + 2R_2'\left(\frac{\partial f}{\partial r_2} + \right. \\ \left. \frac{r_2^2 - r_1^2 + r_3^2}{r_2 r_3}\frac{\partial f}{\partial r_3}\right)\frac{\partial^2\psi}{\partial x_2 \partial x_3} + \left(R_1'' + \frac{2}{r_1}R_1'\right)\frac{\partial\psi}{\partial x_1} + \left(R_2'' + \right. \\ \left. \frac{2}{r^2}R_2'\right)\frac{\partial\psi}{\partial x_2} + \left(\frac{\partial^2 f}{\partial r_1^2} + \frac{\partial^2 f}{\partial r_2^2} + 2\frac{\partial^2 f}{\partial r_3^2} + \frac{r_1^2 - r_2^2 + r_3^2}{r_1 r_3}\frac{\partial^2 f}{\partial r_1 \partial r_3} + \right. \\ \left. \frac{r_2^2 - r_1^2 + r_3^2}{r_2 r_3}\frac{\partial^2 f}{\partial r_2 \partial r_3} + \frac{2}{r_1}\frac{\partial\psi}{\partial r_1} + \frac{2}{r_2}\frac{\partial f}{\partial r_2} + \frac{4}{r_3}\frac{\partial f}{\partial r_3}\right)\frac{\partial\psi}{\partial x_3} + \\ (E - V)\psi = 0, \end{aligned} \tag{4}$$

where primes denote derivatives.

In order that the coördinate system $x_i$ be orthogonal $f$ must satisfy the equations

$$\frac{\partial f}{\partial r_1} + \frac{r_1^2 - r_2^2 + r_3^2}{2r_1 r_3}\frac{\partial f}{\partial r_3} = 0, \qquad \frac{\partial f}{\partial r_2} + \frac{r_2^2 - r_1^2 + r_3^2}{2r_2 r_3}\frac{\partial f}{\partial_{r_3}} = 0 \tag{5}$$

and there is no restriction upon $R_1$ and $R_2$. Eliminating $\partial f/\partial r_3$ from these equations we have

$$\frac{r_2^2 - r_1^2 + r_3^2}{r_2} \frac{\partial f}{\partial r_1} + \frac{r_2^2 - r_1^2 - r_3^2}{r_1} \frac{\partial f}{\partial r_2} = 0.$$

A solution of this equation is

$$f = \frac{r_3^2 - r_1^2 - r_2^2}{r_1 r_2} \tag{6}$$

and any function of this solution is also a solution. Equations (5) are satisfied identically by this value of $f$. Substituting it in equation (4) we obtain

$$R_1'^{\,2} \frac{\partial^2\psi}{\partial x_1^2} + R_2'^{\,2} \frac{\partial^2\psi}{\partial x_2^2} + \frac{r_1^2 + r_2^2}{r_1^2 r_2^2} [2(r_1^2 + r_2^2)r_3^2 - (r_1^2 - r_2^2)^2 - r_3^4] \times \frac{\partial^2\psi}{\partial x_3^2} + \left(R_1'' + \frac{2}{r_1} R_1'\right) \frac{\partial\psi}{\partial x_1} + \left(R_2'' + \frac{2}{r_2} R_2'\right) \frac{\partial\psi}{\partial x_2} + \frac{2}{r_1 r_2} \times \left(\frac{r_2^2 - r_3^2}{r_1^2} + \frac{r_1^2 - r_3^2}{r_2^2} + 2\right) \frac{\partial\psi}{\partial x_3} + (E - V)\psi = 0.$$

By means of

$$x_3 = \frac{r_3^2 - r_1^2 - r_2^2}{r_1 r_2} \tag{7}$$

this equation may be written

$$R_1'^{\,2} \times \left[\frac{\partial^2\psi}{\partial x_1^2} + \left(\frac{R_1''}{R_1'^{\,2}} + \frac{2}{r_1}\frac{1}{R_1'}\right) \frac{\partial\psi}{\partial x_1}\right] + R_2'^{\,2} \left[\frac{\partial^2\psi}{\partial x_2^2} + \left(\frac{R_2''}{R_2'^{\,2}} + \frac{2}{r_2}\frac{1}{R_2'}\right) \frac{\partial\psi}{\partial x_2}\right] + \frac{r_1^2 + r_2^2}{r_1^2 r_2^2} (4 - x_3^2) \left(\frac{\partial^2\psi}{\partial x_3^2} + \frac{2x_3}{x_3^2 - 4} \frac{\partial\psi}{\partial x_3}\right) + (E - V)\psi = 0. \tag{8}$$

For a particular choice of $R_1$ and $R_2$ in (3) we write as the inverses of these equations

$$r_1 = h_1(x_1), \qquad r_2 = h_2(x_2). \tag{9}$$

If we put

$$V = R_1'^{\,2} f_1(x_1) + R_2'^{\,2} f_2(x_2) + \frac{r_1^2 + r_2^2}{r_1^2 r_2^2} (4 - x_3^2) f_3(x_3), \tag{10}$$

where $f_i$ is an arbitrary function of $x_i$, and

$$\psi = X_1 X_2 X_3, \tag{11}$$

where $X_i$ is a function of $x_i$ alone, equation (8) may be written

$$R_1'^{\,2}\left[\frac{X_1''}{X_1} + \left(2\frac{h_1'}{h_1} - \frac{h_1''}{h_1'}\right)\frac{X_1'}{X_1} + \frac{E}{3} - f_1(x_1)\right] +$$
$$R_2'^{\,2}\left[\frac{X_2''}{X_2} + \left(2\frac{h_2'}{h_2} - \frac{h_2''}{h_2'}\right)\frac{X_2'}{X_2} + \frac{E}{3} - f_2(x_2)\right] +$$
$$\frac{r_1^2 + r_2^2}{r_1^2 r_2^2}(4 - x_3^2)\left[\frac{X_3''}{X_3} + \frac{2x_3}{x_3^2 - 4}\frac{X_3'}{X_3} + \frac{E}{3} - f_3(x_3)\right] = 0. \quad (12)$$

Hence the functions $X_1$, $X_2$, $X_3$ such that $\psi = X_1X_2X_3$ is a solution of equation (8) are solutions of the ordinary differential equations obtained by equating to zero the expressions in parentheses in (12).

From (1) and (7) it follows that, if $\theta$ is the angle between the vectors $0P$ and $0\bar{P}$,

$$x_3 = -2\cos\theta.$$

Consequently $V$ as given in (10) may be written

$$V = F_1(r_1) + F_2(r_2) + 4\left(\frac{1}{r_1^2} + \frac{1}{r_2^2}\right)\sin^2\theta f_3(-2\cos\theta), \quad (10a)$$

where $F_1$ and $F_2$ are

$$F_1 = R_1'^{\,2} f_1(R_1), \qquad F_2 = R_2'^{\,2} f_2(R_1)$$

and consequently can be given arbitrary values by suitable choice of $R_1$, $R_2$, $f_1(R_1)$, and $f_2(R_2)$.

Another orthogonal coördinate system $x_i$ is given by

$$x_1 = \frac{r_1 + r_2}{r_3}, \qquad x_2 = \frac{r_1 - r_2}{r_3}, \qquad x_3 = \sqrt{r_1^2 + r_2^2}, \quad (13)$$

as shown by Gronwall.[2] In this coördinate system equation (2) is

$$\frac{r_1^2 + r_2^2}{r_1 r_2 r_3^2}\left[(x_1^2 - 1)\left(\frac{\partial^2\psi}{\partial x_1^2} + \frac{\partial}{\partial x_1}\log\frac{x_1^2 - 1}{x_1^2 + x_2^2}\frac{\partial\psi}{\partial x_1}\right) +\right.$$
$$\left.(1 - x_2^2)\left(\frac{\partial^2\psi}{\partial x_2^2} + \frac{\partial}{\partial x_2}\log\frac{1 - x_2^2}{x_1^2 + x_2^2}\frac{\partial\psi}{\partial x_2}\right)\right] + \frac{\partial^2\psi}{\partial x_3^2} +$$
$$\frac{5}{x_3}\frac{\partial\psi}{\partial x_3} + (E - V)\psi = 0. \quad (14)$$

If we put

$$\psi = \sqrt{\frac{x_1^2 + x_2^2}{(x_1^2 - 1)(1 - x_2^2)}}\,\varphi,$$

the equation in $\varphi$ is

$$\frac{r_1^2 + r_2^2}{r_1 r_2 r_3^3}\left[(x_1^2 - 1)\frac{\partial^2\varphi}{\partial x_1^2} + \frac{1}{x_1^2 - 1}\varphi + (1 - x_2^2)\frac{\partial^2\varphi}{\partial x_2^2} + \frac{1}{1 - x_2^2}\varphi\right] + \frac{\partial^2\varphi}{\partial x_3^2} + \frac{5}{x_3}\frac{\partial\varphi}{\partial x_3} + \left(E + \frac{2}{x_3^2} - V\right)\varphi = 0.$$

If then we put

$$V = \frac{r_1^2 + r_2^2}{r_1 r_2 r_3^2}\left[f_1\left(\frac{r_1 + r_2}{r_3}\right) + f_2\left(\frac{r_1 - r_2}{r_3}\right)\right] + f_3(r_1^2 + r_2^2), \tag{15}$$

where the functions $f_i$ are arbitrary, a solution of (14) is given by

$$\psi = \sqrt{\frac{x_1^2 + x_2^2}{(x_1^2 - 1)(1 - x_2^2)}}\, X_1 X_2 X_3,$$

where

$$(x_1^2 - 1)X_1'' + \left(\frac{1}{x_1^2 - 1} - f_1(x_1)\right) X_1 = 0,$$

$$(1 - x_2^2)X_2'' + \left(\frac{1}{1 - x_2^2} - f_2(x_2)\right) X_2 = 0,$$

$$X_3'' + \frac{5}{x_3} X_3' + \left(E + \frac{3}{x_3} - f(x_3^2)\right) X_3 = 0.$$

[1] Hylleraas, E. A., Neue Berechnung der Energie des Heliums im grundzustande, sowie des tiefsten Terms von Ortho-Helium, *Z. Physik.*, **54**, 347–366 (1929).

[2] Gronwall, T. H., A special conformally euclidean space of three dimensions occurring in wave mechanics, *Ann. Math.*, Ser. 2, **34**, 279–293 (1932).

---

## *ERRATUM*

In Table 2 of the article "Aggregation Phenomena in Polyvinyl Alcohol-Acetate Solutions" on page 367 of the July issue, 1949, the headings of columns two and four should be interchanged. F. F. NORD

PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES

Volume 35 September 15, 1949 Number 9

---

# ON THE STABILITY OF THE $S_8$ MOLECULE AND THE STRUCTURE OF FIBROUS SULFUR

By Linus Pauling

California Institute of Technology, Pasadena*

Communicated June 15, 1949

Sulfur is unique among the elements in consisting in its standard state of octatomic molecules. It is interesting that the stability of the $S_8$ molecule can be understood by consideration of the forces determining the orientation around the sulfur-sulfur single bond.

Let us consider a molecule such as $H_2S_2$, in which each sulfur atom forms a single bond with the other sulfur atom and a single bond with another atom (hydrogen), and has in addition two unshared electron pairs in its outer shell. The bonds formed by the sulfur atoms involve mainly $3p$ orbitals, and one of the two unshared outer electron pairs occupies the third $3p$ orbital, the other unshared pair occupying essentially the $4s$ orbital. (A rough theoretical treatment which will be communicated later indicates that the bond orbitals have only about 1% $s$ character.) The dihedral angle between the planes H—S—S and S—S—H is determined principally by the repulsion of $p\pi$ electrons. The repulsion between $p\pi$ electrons is at a maximum when the $p\pi$ orbitals of the two sulfur atoms are in the same plane, and falls to zero when these orbitals are in orthogonal planes. The energy of interaction of an unshared pair of $p\pi$ electrons on one atom and the similar unshared pair on the other atom in a coplanar $p\pi$ orbital is equal to $-2\alpha$ (the negative sign corresponding to repulsion, because $\alpha$, the exchange integral occurring in the electron-pair bond theory, is itself negative), that between an unshared pair on one atom and a single bonding electron on another is $-\alpha$, and that between two bonding electrons is $-{}^1/_2\alpha$. Hence the total interaction energy of the $p\pi$ electrons in $H_2S_2$ varies from $-{}^5/_2\alpha$ for a planar configuration of the molecule (dihedral angle $\gamma$ equal to 0° or 180°) to a minimum of $-2\alpha$ for $\gamma = 90°$. It is then to be predicted that the molecule $H_2S_2$ has a skew configuration, with the dihedral angle $\gamma$ equal to approximately 90°. The configuration of this molecule has not been precisely determined, although it is known not to be planar,[1] but the similar molecule hydrogen peroxide, $H_2O_2$, has been shown to have $\gamma$ equal to approximately 106° in the crystal formed by hydrogen peroxide and urea[2] and a similar value in the pure substance.[3] (An explanation essentially equivalent to that given above was advanced by Penney and Sutherland to account for the skew configuration of hydrogen peroxide.[4]) Moreover, Guthrie has found, from a careful electron-diffraction investigation of sulfur monochloride, $S_2Cl_2$, that the dihedral angle in this molecule is equal to $92° \pm 12°$,[5] and the value of $\gamma$ in dimethyltrisulfide has been found[6] to be approximately 106°. We accordingly conclude that the normal value of the dihedral angle $\gamma$ for the sulfur-sulfur bond is in the neighborhood of 100°. The increase over the value 90° given by the simple theoretical treatment may be due to van der Waals repulsion of the non-bonded atoms and to other interactions, such as the electrostatic interactions of multipoles discussed recently by Lassettre and Dean.[7]

TABLE 1

VALUES OF THE DIHEDRAL ANGLE IN STAGGERED RING MOLECULES

| NUMBER OF ATOMS IN RING | BOND ANGLE | | | | | | |
|---|---|---|---|---|---|---|---|
| | 90° | 95° | 100° | 105° | 110° | 115° | 120° |
| 4 | 0° | ..... | ..... | ..... | ..... | ..... | ..... |
| 6 | 90° | 84°31′ | 77°52′ | 69°34′ | 58°41′ | 42°13′ | 0° |
| 8 | 114°28′ | 111°0′ | 106°50′ | 102°6′ | 96°18′ | 89°10′ | 80°7′ |
| 10 | 128°10′ | 125°34′ | 122°32′ | 118°59′ | 114°48′ | 109°47′ | 103°39′ |
| 12 | 137°4′ | 134°57′ | 132°31′ | 129°41′ | 126°21′ | 122°24′ | 117°39′ |

In the molecule $S_8$, which has the staggered configuration, the S—S—S bond angle $\beta$ has been found to be equal to 105° both in the orthorhombic crystal[8] and in the vapor.[9] The dihedral angle $\gamma$ is not independently variable, but is determined by $\beta$; its value is 102°. This value is, according to the foregoing argument, close to the normal value, which minimizes the energy of the X—S—S—X group; and it is our thesis that the $S_8$ molecule is stable (relative to $S_x$, with $x \neq 8$) because its configuration leads to this value.

The relation between $\gamma$ and $\beta$ for symmetrical staggered rings (with point-group symmetry $D_{nh}$) containing $n$ atoms ($n$ even) is

$$\sin^2\frac{\gamma}{2} = \frac{\cos(2\pi/n) + \cos\beta}{1 + \cos\beta}. \qquad (1)$$

In table 1 there are given values of $\gamma$ for values of $\beta$ between 90° and 120° and for $n$ between 4 and 12. It is seen that for $\beta = 105°$, the value observed for sulfur, the minimum deviation from the normal value of $\gamma$, which we may take as 100°, occurs at $n = 8$, and amounts to about 2° as compared with 19° for $n = 10$ and 30° for $n = 6$. Since the staggered configurations of the ring provide better approximation to the normal value of $\gamma$ than do other configurations, this calculation gives an explanation of the stability of $S_8$ relative to other sulfur molecules. At higher temperatures sulfur vapor contains $S_6$ and $S_2$ as well as $S_8$. There is little doubt that the $S_6$ molecule is a staggered hexagonal ring. The difference in heat content of $S_6$ and $S_8$ has been evaluated from thermochemical data by Bichowsky and Rossini,[10] and found to be equal to 1.25 kcal./mole per sulfur-sulfur bond. The interaction energy of the $p\pi$ electrons is according to simple theory proportional to $\sin^2 \gamma$. On this assumption, and the assumptions that the bond angle $\beta$ is the same in $S_6$ as in $S_8$ and that the normal value of $\gamma$ is 100°, $\alpha$ can be evaluated as $-9.8$ kcal./mole. This value for $\alpha$ can be considered to be only a very rough one, inasmuch as other structural factors also without doubt contribute to the difference in energy of $S_6$ and $S_8$, but it is worth pointing out that the value is not unreasonable.

The standard state of selenium is a hexagonal crystalline state in which there are infinite spiral molecules of selenium atoms, with Se—Se—Se bond angle $\beta = 105°$. There also exists a less stable red crystalline state of selenium containing $Se_8$ molecules, and the vapor of selenium consists of $Se_8$ molecules, $Se_6$ molecules, and, at higher temperatures, $Se_2$ molecules. Bichowsky and Rossini give for the difference in energy of $Se_6$ and $Se_8$ the value 1.07 kcal./mole of Se—Se bonds, which corresponds to $-8.4$ kcal./mole for the resonance integral $\alpha$.

The spiral molecules in the hexagonal crystals of selenium, and also those representing the standard state of tellurium, have a threefold axis. The relation between the dihedral angle $\gamma$ and the bond angle $\beta$ for a threefold spiral molecule is $\sin^2 \gamma = {}^1/_2\,(1 - 2 \cos \beta)/(1 - \cos \beta)$, corresponding for the values 90°, 95°, 100°, 105°, 110°, 115° and 120° for $\beta$ to the values 90°, 94°36′, 98°31′, 101°52′, 104°46′, 107°17′ and 109°28′ for $\gamma$. It is seen that for $\beta = 105°$, as observed for sulfur and selenium, the value of $\gamma$ is nearly the same for the threefold spiral as for the 8-membered staggered ring. The greater stability of the crystals containing the threefold spiral for selenium and tellurium is presumably due to the strong interaction between neighboring molecules which is made evident in the semi-metallic properties of the crystals. For tellurium, with bond angle 102°, the value of $\gamma$ is 100°.

The failure of oxygen to form a molecule such as $O_8$ is of course due to the greater stability of the multiple bonds in the $O_2$ molecule; this greater stability of multiple bonds over single bonds is characteristic of the first-row elements.

It is interesting that fibrous sulfur, made by stretching the rubbery form of sulfur obtained on cooling liquid sulfur-$\mu$, seems not to contain spirals with a threefold axis, like the spirals in selenium and tellurium, but to have another structure. Meyer and Go[11] have reported that the unit of structrue of fibrous sulfur is monoclinic, with $a_0$ = 26.4 Å, $b_0$ = 9.26 Å, $c_0$ = 12.32 Å and $\beta$ = 79°15′, and that it contains 112 atoms. The $b$ axis is the fiber axis. These authors suggested that there are 14 chains with a repeating unit of eight atoms each in the unit, and proposed two alternative planar configurations for the chains. These planar configurations are, however, to be rejected, because they make the S—S—S—S dihedral angle equal to 0°, which corresponds to the minimum stability. It might indeed be expected that the sulfur atoms are arranged in a simple threefold spiral, with two turns in the unit distance $b_0$, as was tentatively suggested by Huggins.[12] This, however, leads to the value 109° for the S—S—S bond angle, which is unsatisfactory, and, moreover, the reported number of atoms in the unit is not a multiple of 6.

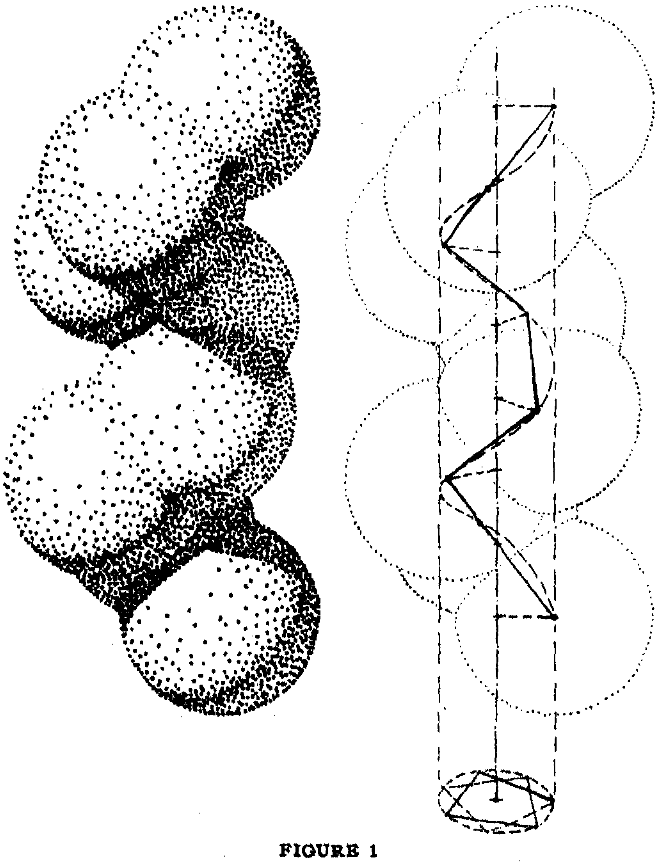


FIGURE 1

The seven-atom two-turn spiral proposed as the repeating unit in the chains of fibrous sulfur.

A conclusion about the probable structure can be reached from the assumption that the S—S distance and the S—S—S bond angle have the same values (2.08 Å and 105°) as in $S_8$. It is found that no spiral making a single turn in the identity distance $b_0$ is acceptable. A spiral making two turns in $b_0$ would be predicted, with the above parameters, to have $b_0 = 8.89$ Å for six atoms, 9.12 Å for seven and 8.49 Å for eight. Only the second of these is close enough to the observed value 9.26 Å to be indicated as correct; indeed, increasing the bond angle to 106° leads for the seven-atom two-turn spiral to exactly $b_0 = 9.26$ Å. The dihedral angle for this spiral is 76°, which is intermediate between the value for $S_8$ and that for $S_6$, and, through its derivation from the value corresponding to maximum stability, is compatible with the ease of conversion of fibrous sulfur into orthorhombic sulfur. The seven-atom chain unit explains the unusual occurrence of a multiple of seven atoms (112) in the unit of structure; the packing of chains is such as to lead to sixteen chains, parallel to $b$, per unit. Moreover, the fact that $b_0$ is 9.26 Å rather than approximately half so large is explained by the seven-atom spiral, which repeats only after seven atoms, whereas the six-atom and eight-atom spirals (with two turns) would repeat after a single turn. It is accordingly very likely that the structural unit of fibrous sulfur is the seven-atom two-turn spiral, shown in figure 1. The geometrical properties of the spiral are such that the distances $b_0 = 9.26 \pm 0.05$ Å and S—S $= 2.08 \pm 0.02$ Å determine the bond angle to be $106° \pm 1/2°$.

---

[*] Contribution No. 1296 from the Gates and Crellin Laboratories of Chemistry.

[1] Badger, R. M., private communication.

[2] Lu, C.-S., Hughes, E. W., and Giguère, P. A., *J. Am. Chem. Soc.*, **63**, 1507 (1941).

[3] Zumwalt, L. R., and Giguère, P. A., *J. Chem. Phys.*, **9**, 458 (1941).

[4] Penney, W. G., and Sutherland, G. B. B. M., *Ibid.*, **2**, 492 (1934).

[5] Guthrie, G. B., Jr., Dissertation, California Institute of Technology, 1949.

[6] Donohue, J., and Schomaker, V., *J. Chem. Phys.*, **16**, 92 (1948).

[7] Lassettre, E. N., and Dean, L. B., Jr., *Ibid.*, **17**, 317 (1949).

[8] Warren, B. E., and Burwell, J. T., *Ibid.*, **3**, 6 (1935).

[9] Lu, C.-S., and Donohue, J., *J. Am. Chem. Soc.*, **66**, 818 (1944).

[10] Bichowsky, F. R., and Rossini, F. D., "The Thermochemistry of the Chemical Substances," Reinhold Publishing Corporation, New York, 1936.

[11] Meyer, K. H., and Go, Y., *Helv. chim. Acta*, **17**, 1081 (1934).

[12] Huggins, M. L., *J. Chem. Phys.*, **13**, 37 (1945).

# EVIDENCE FOR A TRYPTOPHANE CYCLE IN NEUROSPORA*

BY FRANCIS A. HASKINS AND HERSCHEL K. MITCHELL

THE KERCKHOFF LABORATORIES OF BIOLOGY, CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

Communicated by G. W. Beadle, July 15, 1949

The work of Snell[1] on lactic acid bacteria provided the first evidence that anthranilic acid may act as a precursor to tryptophane *in vivo*. Since the strains of bacteria which utilize anthranilic acid in the place of tryptophane also use indole, it was not possible to establish the order in which these two precursors occurred. In 1943 Tatum, Bonner and Beadle[2] found that anthranilic acid was accumulated in the culture medium by a *Neurospora crassa* mutant, strain 10575, which required either tryptophane or indole for growth. It was therefore indicated that anthranilic acid is a precursor to indole in Neurospora.

It has recently been observed in this laboratory that when a medium containing tryptophane is inoculated with wild type or any one of a number of tryptophane-utilizing strains of Neurospora, a blue fluorescence, resembling that of an anthranilic acid solution, is produced. The fluorescence is detectable within a few hours after inoculation, reaches a maximum intensity at 24 to 30 hours and then diminishes until, at 72 hours, the medium displays essentially no fluorescence except in the case of the mutants which are able to use tryptophane but cannot utilize anthranilic acid. Earlier work has already demonstrated that anthranilic acid is formed by the degradation of tryptophane by bacteria[3, 4] and in the rat and other animals.[5] Thus the appearance and subsequent disappearance of fluorescence in Neurospora medium suggested that in Neurospora some of the reactions by which tryptophane is degraded and synthesized may constitute a metabolic cycle. The present experimental evidence supports this conclusion.

An earlier communication[6] from this laboratory postulated 3-hydroxykynurenine as an intermediate in the conversion of tryptophane to nicotinic acid in Neurospora. Through the generosity of Dr. W. Weidel, of the Kaiser Wilhelm Institut für Biochemie, a sample of this compound was made available to the authors and it became possible to establish its relation to the tryptophane cycle in Neurospora. The substance was isolated by Weidel in 1949 from larvae of *Calliphora erythrocephala* as the $cn^+$ substance (precursor of eye pigments in insects).

*Experimental.—Mutant Strains.*—The qualitative growth responses of the various Neurospora strains considered in this investigation are listed in table 1. Previous descriptions have been made of strains E5212,[7] C86,[7, 8] 10575,[2] C83[8, 9] and 4540.[10]

*Disappearance of Tryptophane from the Culture Medium.*—Erlenmeyer flasks (125 ml.) containing 20 ml. of minimal medium[11] supplemented with 400 γ of L-tryptophane were inoculated with conidia from Neurospora strain B1312 and were incubated at 25°C. Absorption spectra of the medium before inoculation and at 8-hour intervals thereafter were determined with a Beckman spectrophotometer. These spectra are shown in Figure 1. Dry weights of mold per flask at the different times were as follows: 16 hours, 0.6 mg.; 24 hours, 3.0 mg.; 32 hours, 9.1 mg. The subsequent dry weight attained at 96 hours was 55 mg. On the basis of the spectra at 280 mμ (the absorption maximum for the indole ring of tryptophane), it is obvious that at least 100 γ of the tryptophane are gone when the growth of mold is only 0.6 mg. By the time 3 mg. of growth has been produced, at least half of the tryptophane has disappeared, and the shape of the absorption spectrum suggests that actually much more than half is gone.

In order to eliminate the possibility that the tryptophane was not simply absorbed from the medium by the small amount of mold present in the early stages of growth, the mold was removed from 32- and 96-hour cultures, hydrolyzed with 1 *N* NaOH, and the tryptophane content was determined by bioassay with *S. lactis R.*[1] A control sample of tryptophane was subjected to the same treatment. The tryptophane content of the 32-hour culture on a dry-weight basis was found to be 0.3% and that of the 96-hour culture 0.2%. After 32 hours the mold therefore contains approximately 27 γ of tryptophane as opposed to a disappearance of at least 200 γ. Thus it is clearly demonstrated that the tryptophane has not been taken up as such by the mold, but has actually been converted to other compounds.

TABLE 1

UTILIZATION OF VARIOUS COMPOUNDS BY *Neurospora* MUTANTS

| STRAIN | TYROSINE AND CINNAMIC ACID | PHENYLALANINE | ANTHRANILIC ACID | INDOLE | TRYPTOPHANE | KYNURENINE | 3-HYDROXYKYNURENINE | 3-HYDROXYANTHRANILIC ACID | NICOTINIC ACID |
|---|---|---|---|---|---|---|---|---|---|
| E5212 | − | + | − | − | − | − | − | − | − |
| C86 | + | + | + | + | + | + | + | + | + |
| B1312 | − | − | + | + | + | + | − | − | − |
| 10575 | − | − | − | + | + | − | − | − | − |
| C83 | − | − | − | − | + | − | − | − | − |
| E5029 | − | − | − | − | − | − | + | + | + |
| 4540 | − | − | − | − | − | − | − | − | + |

*Products of Tryptophane Degradation.*—Flasks of medium supplemented with 400 γ of L-tryptophane per 20 ml. were inoculated with conidia from wild-type strain 8a and from mutants B1312, C86 and 10575, respectively. After 32 hours' incubation at 25° the mold was removed from the flasks, the medium was filtered, the pH was adjusted to 2.5 and the medium was extracted four times with an equal volume of ether. The ether was evaporated on a warm water bath. With each strain the absorption spectrum of the material in the ether phase was found to be nearly the same as the absorption spectrum of the unextracted medium. Measurements of fluorescence using a Coleman photofluorometer disclosed that approximately 80% of the fluorescence originally present in the unextracted medium went to the ether phase. The residual aqueous phase had an absorption spectrum corresponding to a maximum of 50 γ of tryptophane. Bioassays, using strains of C86 and B1312, showed that from the original 400 γ of tryptophane there remained in the medium after 32 hours' growth the biological equivalent of approximately 250 γ. Of this, approximately 180 γ were extracted with ether. With the unextracted medium, strains 10575 and C83 which do not use anthranilic acid gave growth which was equivalent to that obtained on about 50 γ of tryptophane, but neither of these strains grew on the ether extract. All four of the strains tested gave the same general qualitative and quantitative results as determined by measurements of absorption spectra, fluorescence and biological activity.

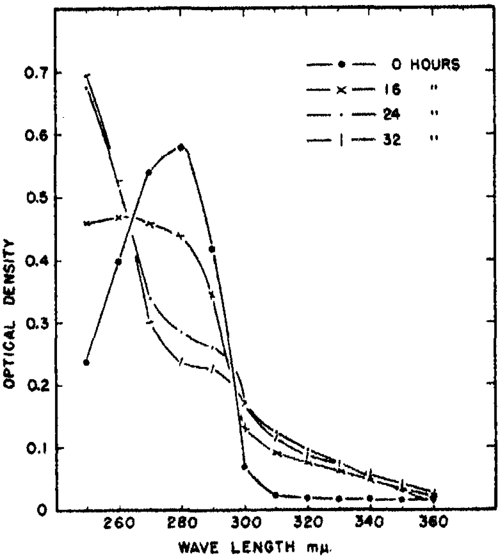


FIGURE 1

Absorption spectra of *Neurospora* medium at various times after inoculation.

Filter paper strip chromatograms were run on the ether extract with water-saturated butanol as the solvent. After development, the chromatograms were examined under ultra-violet light to detect fluorescent areas, and were then cut into sections, the sections were extracted with 0.1 *M* $KH_2PO_4$ (pH 3.5), and the biological activities of the resulting solutions were determined using B1312 for the assays. It was found that approximately 60% of the activity of the ether phase could be accounted for by the activity of a chromatogram band which resembled an anthranilic acid band in its position and fluorescence.

In order to obtain enough material for isolation of some of the degradation products, 8 liters of culture medium containing 8 g. of L-tryptophane was inoculated with a heavy inoculum of conidia from wild-type strain 8a. The culture was incubated with forced aeration in a 35° water bath for 19 hours. The mycelium was then filtered off, the pH of the filtrate was adjusted to 2.5, the filtrate was divided into two portions and each was extracted with seven 500-ml. portions of ether. Following evaporation of the ether, the residual solution was absorbed on 29 sheets of filter paper and these sheets were incorporated into a chromatopile[12] consisting of 600 sheets of 9-cm. Whatman No. 1 filter paper. Water-saturated butanol was used as the developer. After 15 hours, every twentieth sheet was removed from the pile and the location of anthranilic acid was determined by absorption spectra of extracts of the sample sheets. The desired section was then removed from the chromatopile, dried in a vacuum desiccator over $CaCl_2$ and extracted with ether in a Soxhlet. The residue from evaporation of the ether was sublimed at 100° *in vacuo* (20 microns Hg). The sublimate was crystallized from water and further purified by resublimation in a temperature gradient at 20 microns pressure. The material obtained in this way melted at 143–144° (cor.). A known sample of anthranilic acid melted at 144–145° (cor.) and a mixture of the two melted at 144–145° (cor.). The neutralization equivalent of the isolated compound was 137 (calculated for anthranilic acid—137). The absorption spectrum, fluorescence and biological activity were identical with that of anthranilic acid, furnishing further proof that the isolated compound was indeed anthranilic acid.

Chromatographic examination of the products of tryptophane degradation by germinating Neurospora has demonstrated the existence of at least two biologically active substances besides anthranilic acid. In addition bio-assay with *L. arabinosus* has shown the production of 4 $\gamma$ of nicotinic acid from 400 $\gamma$ of tryptophane in 32 hours.

A number of compounds other than tryptophane were tested to determine whether they were able to give rise to fluorescent degradation products in the medium. Strain C86 was used as the test strain. It was found that strain C86 produced the typical blue fluorescence in the medium when supplied with phenylalanine, indole or kynurenine. No appreciable fluorescence was produced when the medium was supplemented with trans-cinnamic acid, tyrosine, 3-hydroxykynurenine or nicotinamide. A test on 3-hydroxyanthranilic acid appeared also negative but inasmuch as the compound itself had some blue fluorescence the test was not critical. In the cases where fluorescence was produced it was not permanent but reached a maximum and then disappeared, as did the fluorescence produced when C86 was supplied with tryptophane.

*Discussion.*—The evidence reported here establishes the fact that tryptophane disappears rapidly from culture medium which has been inoculated with wild-type Neurospora strain 8a or mutant strains C86, B1312 or 10575. A large part of the tryptophane is gone before appreciable growth of the mold has occurred. It is also established that anthranilic acid is one of the products of tryptophane breakdown in Neurospora.

The accumulated evidence points clearly to the existence in Neurospora of a metabolic cycle involving, among other compounds, tryptophane and anthranilic acid. A schematic representation of such a cycle is shown in figure 2. Since strain B1312 is able to utilize anthranilic acid or kynurenine, but not 3-hydroxykynurenine, it is apparent that kynurenine is an intermediate in the formation of anthranilic acid from tryptophane and that 3-hydroxykynurenine is not. In addition the mold will produce blue fluorescence in the medium when supplied with kynurenine, but not when supplied with the hydroxy compound. The reactions by which anthranilic acid is formed from tryptophane must include at least one essentially irreversible step, for strains 10575 and C83 cannot utilize anthranilic acid although they

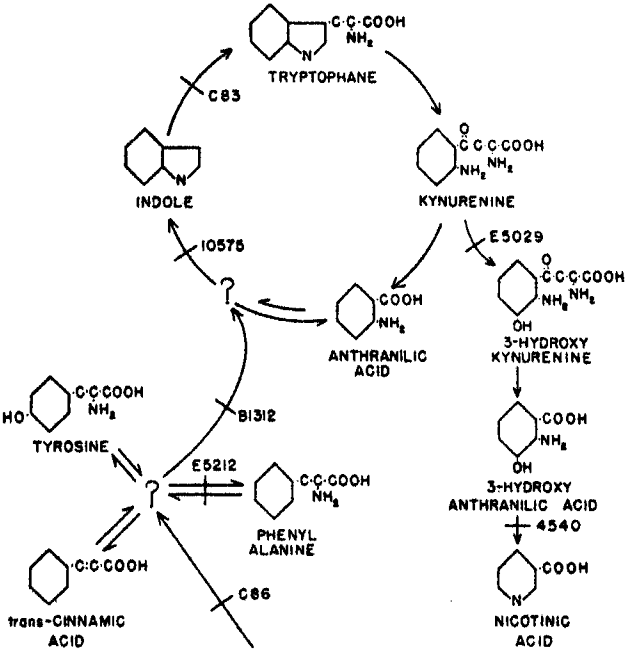


FIGURE 2

Schematic representation of a tryptophane cycle in *Neurospora.*

are able to make it from tryptophane. This is in agreement with the finding by Kotake[5] that rats can make anthranilic acid from tryptophane, but are evidently unable to use anthranilic acid to satisfy their tryptophane requirement.

As previously reported[13] the carboxyl carbon atom of anthranilic acid is lost as $CO_2$ during metabolism of this substance by Neurospora. It seems most probable that this occurs in the conversion of anthranilic acid to indole since Heidelberger has traced the corresponding carbon of the indole ring of tryptophane to kynurenine and nicotinic acid in rats.[14, 15, 16] For this reason anthranilic acid has tentatively been excluded as an intermediate between phenylalanine and indole in the diagram in figure 2. The accumulation of anthranilic acid by strain 10575[2] is not sufficient reason for placing anthranilic acid immediately before indole since strain C83 also accumulates this substance. It seems more likely that the accumulation of anthranilic acid is due to a reversible reaction between this compound and an unknown intermediate with anthranilic acid formation being favored as shown in the diagram. The accumulation of anthranilic acid during tryptophane degradation by wild type is in agreement with this conclusion. This problem requires further elucidation before the position of anthranilic acid in the cycle can be established with certainty. The number or nature of all the intermediates in the cycle is not known, but as stated above, there are at least three biologically active substances other than anthranilic acid produced as the result of tryptophane degradation. The task of isolating and identifying these intermediates and side products is in progress.

It is clear from the experimental evidence that at least one-fourth of the tryptophane degraded by the mold is converted to substances lacking biological activity for the tryptophane series of mutants. It is not yet clear, however, whether or not these compounds or the cycle reactions play essential roles in the metabolism of the mold. It is to be noted that the mutants which cannot carry out reactions within the cycle (10575 and C83) grow poorly, as compared to all the other mutants of the series, even with an excess of tryptophane. This fact suggests a special metabolic significance to turnover in the cycle.

*Summary.*—1. Tryptophane is shown to disappear rapidly from culture medium which is inoculated with any one of the Neurospora strains tested. One of the products of tryptophane breakdown has been isolated and identified as anthranilic acid.

2. Evidence is presented which suggests that in Neurospora tryptophane, anthranilic acid and several other compounds are involved in a metabolic cycle.

3. The evidence demonstrates that 3-hydroxykynurenine is an intermediate in tryptophane metabolism in Neurospora, but indicates that it is not directly involved in the cycle.

* This work was supported by funds from the Rockefeller Foundation and by funds from the Atomic Energy Commission administered through contract with the Office of Naval Research, United States Navy, Contract N6 onr-244 Task order V.

[1] Snell, E. E., *Archiv. Biochem.*, **2**, 389 (1943).

[2] Tatum, E. L., Bonner, D., and Beadle, G. W., *Ibid.*, **3**, 477 (1943).

[3] Sasaki, T., *J. Biochem. (Japan)*, **2**, 251 (1923); *Chemical Abstracts*, **17**, 1817 (1923).

[4] Kotake, Y., and Otani, S., *Z. Physiol. Chem.*, **214**, 1 (1933).

[5] Kotake, Y., *J. Chem. Soc. Japan*, **60**, 632 (1939); *Chemical Abstracts*, **36**, 4141 (1942).

[6] Mitchell, H. K., and Nyc, J. F., PROC. NAT. ACAD. SCI., **34**, 1 (1948).

[7] Nyc, J. F., Haskins, F. A., and Mitchell, H. K., *Archiv. Biochem.* (in press) (1949).

[8] Lein, J., Mitchell, H. K., Houlahan, M. B., PROC. NAT. ACAD. SCI., **34**, 435 (1948).

[9] Mitchell, H. K., and Lein, J., *J. Biol. Chem.*, **175**, 481 (1948).

[10] Bonner, D., and Beadle, G. W., *Archiv. Biochem.*, **11**, 319 (1946).

[11] Beadle, G. W., and Tatum, E. L., *Am. J. Bot.*, **32**, 678 (1945).

[12] Mitchell, H. K., and Haskins, F. A., *Science* (in press) (1949).

[13] Nyc, J. F., Mitchell, H. K., Leifer, E., and Langham, W. H., *J. Biol. Chem.*, **179**, 783 (1949).

[14] Heidelberger, C., *Ibid.*, **179**, 139 (1949).

[15] Heidelberger, C., Gullberg, M. E., Morgan, A. F., Lepkovsky, S., *Ibid.*, **179**, 143 (1949).

[16] Heidelberger, C., Abraham, E. P., and Lepkovsky, S., *Ibid.*, **179**, 151 (1949).

---

# THE CORRELATION IN SHAPE AND SIZE BETWEEN EPIDERMAL AND SUBEPIDERMAL CELLS

BY FREDERIC T. LEWIS

DEPARTMENT OF ANATOMY, HARVARD MEDICAL SCHOOL, BOSTON, MASSACHUSETTS

Communicated by G. B. Wislocki, July 21, 1949

In these PROCEEDINGS in 1945, Matzke[1] published his remarkable tabulation showing that in a foam of bubbles of equal volume 400 of those at the periphery had an average of 10.99 facets. An average close to 11 was not unexpected, but the precision of his result is impressive. The general shape which such bubbles, or comparable simple vegetable cells, could take uniformly has recently been figured by the writer,[2] using a conventional model. Each peripheral cell has a free surface which is hexagonal; six lateral surfaces, four of which are pentagons, and the other two, on opposite sides, are quadrilateral; and four basal facets, two quadrilateral and two hexagonal. It is immediately apparent that this shape does not depend on uniformity in volume with the cells beneath, since all the conditions may be met by tall prismatic 11-hedral cells of relatively large volume, or by flat hexagonal plates like a thin tiling. But the area of the peripheral or epidermal cell in a tangential section of a cylindrical stem should be the same as that of the underlying or subepidermal cell. Then this pattern may be recognized though, through cell division, the 11-hedron has become an average of diverse shapes and sizes, or, as in bubbles, it is imperfectly presented through casual irregularities.

Frequently the epidermal cells are smaller, or sometimes larger, in tangential section than the cells immediately beneath. This paper deals with the effect of such differences upon the typical pattern. The problem was first encountered by the writer in 1928,[3] in studying the epidermis of the cucumber, and was tentatively solved by use of the diagram here reproduced as figure 1. When the average hexagonal cell of the epidermis has the same transverse dimensions as the cells below, the interrelation of the bases of the epidermal cells and the tops of the underlying cells appears as in *A*. The subepidermal cells are lightly outlined in the figure. It may be assumed, as is often the case, that cells of the *second* layer beneath the epidermis are of the same size as those in the first, so that the subepidermal layer rests upon one which duplicates it. Then the subepidermal cells, if uniform, are 14-hedra, hexagonal in cross-section. They have 6 lateral facets and end in 4 facets above and below. When a cell of this size and description becomes an epidermal cell, a single facet—the hexagonal free surface—replaces one of its 4-faceted ends and makes it 11-hedral. Should the upper and the lower ends each be replaced by a single free surface, the 14-hedron is reduced to an 8-hedron, hexagonal above and below, with its 6 lateral surfaces all quadrilateral.

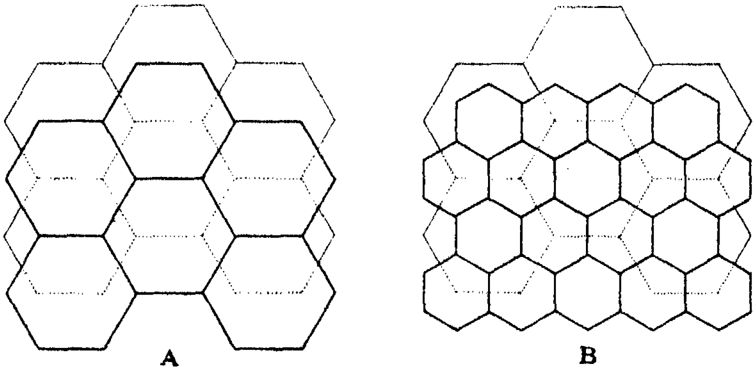


FIGURE 1

*A*.—Layer of regular hexagons superimposed on a layer of regular hexagons of the same area so that the resulting pattern is a projection of tetrakaidecahedra. *B*.—Layer of regular hexagons superimposed on regular hexagons of three times their area. The underlying cells become 17-hedra: the overlying cells are one 8-hedron for two 10-hedra. (*Anat. Rec.*, **38**, 352 (1928)).

An elegant demonstration of the 8-hedral form of cells with free surfaces above and below has recently been provided by Miss Flint[4] in her study of cell shape in the aquatic plant *Utricularia inflata*. Its floats (Schwimmorgane), as figured by Goebel, are bladder-like formations containing an axial strand from which 6 plates of cells, one cell thick, radiate to an enclosing epidermis, also one cell thick, thus dividing the bladder into air-filled compartments. Though the outlines of the septal cells on surface view are wavy, suggesting a jig-saw puzzle, whereas those of the epidermis are "regular," they both present the common range for polygonal cell shapes, from quadrilaterals to nonagons, inclusive, and their average number of sides is nearly 6. If one accepts the geometrical conclusion that, apart from vertices of 4 or more rays and the requisite deduction of 12 sides for a mosaic that completely encloses a space, the average number of sides for such cells must be precisely 6, then the averages recorded are due to the chance of selection. For the free surfaces of 200 septal cells Miss Flint found an average of 5.875 sides, and for 200 epidermal cells 5.825. Every cell had a corresponding number of lateral quadrilateral facets, averaging therefore, instead of 8.0, 7.875 and 7.825 facets, respectively. Large air spaces in the midribs of certain leaves offered a third source of such cells; 200 of them had free surfaces averaging 5.905 sides and a corresponding average of 7.905 facets.

In the cucumber the subepidermal cells, from a few measurements of their transverse diameters, were estimated to have cross-sectional areas about three times as large as the epidermal. Subsequent planimeter measurement of a patch of 100 epidermal cells in that specimen, and of 100 related subepidermal cells in large groups, showed their relative areas more closely as 1:2.66. Since the difference in size develops gradually, there is no expectation that the actual ratios should be the simple fractions used in this study.

Showing precisely the 1:3 proportion, the diagram figure 1, *B*, was presented as "a theoretical abstraction." Its epidermal cells occur in two forms—10-hedra, with 3-faceted bases, surrounding 8-hedra with single-faceted bases, placed over 17-hedral subepidermal cells. Although central 8-hedral epidermal cells certainly occur in this epithelium, they are not sufficiently numerous to form one-third of all its cells as in the pattern figured. This arrangement would reduce the average number of epidermal facets from 11 to 9 $^1/_3$.

A further criticism of the diagram is that the orientation of the hexagons is prosenchymal in one layer and parenchymal in the other. Since these orientations depend on the predominant direction of growth, in girth or in length, respectively, they usually accord in adjacent layers. In the cucumber, expanding comparably in both directions, the orientation is similar, though confused, in these two strata.

When a network of regular hexagons is placed over a network of larger regular hexagons in the same orientation, a complex pattern results which is far more suggestive of the cucumber tissue. By making 7 apothems of the small hexagons equal to 4 apothems of the large, the area of the small will be nearly one-third (32.65 per cent) of the large, and a pattern formed by 49 small hexagons over 16 large ones may be repeated indefinitely. Seven of the 49 small hexagons will provide 8-hedral central or enclosed cells with single-faceted bases (like those in Fig. 1, *B*) and two more will be added after the elimination of coincident edges. The average number of facets for the epidermal polyhedra becomes 9.47, and for the subepidermal, 17.56.

The next tissue to be studied in this respect was fortunately the simpler epidermis of the stem of *Tradescantia*.[5] Its hexagonal cells so far approach rectangular forms that they may be treated as such while recognizing the full significance of certain rounded corners and edges. The epidermal and subepidermal cells, primarily of similar dimensions, increase equally in width, and would also in length except that transverse divisions are more frequent in the epidermis. Hence, though the tissue comes to consist of cells elongated lengthwise of the stem, those of the epidermis are only one-third of the length and area of the subepidermal. An average of $9\,{}^{2}/_{3}$ facets for the mature epidermal cells has been reported, and of 18 facets for those immediately beneath. This contrasts with the epidermal average of $9\,{}^{1}/_{3}$ and subepidermal of 17 facets originally assigned to the cucumber, and also with the revision of $9\,{}^{1}/_{2}$ and $17\,{}^{1}/_{2}$ respectively, now adopted to provide for similar orientation. The significance of these differences may be explained as follows.

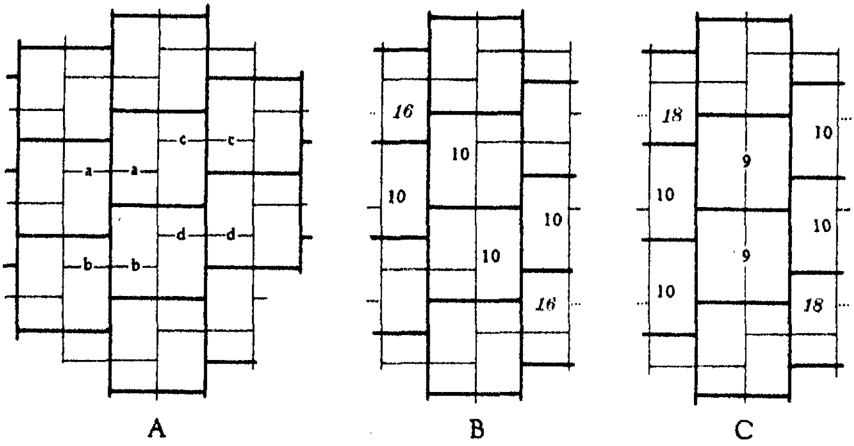


FIGURE 2

*A*.—Rectangular hexagons of the same size superimposed as in Fig. 1, *A*. *B*.—Middle portion of *A* with septa *aa* and *dd* removed. Roman figures indicate the facets of the upper "cells" (always 7 plus their basal facets): italics are the ultimate facet numbers for the cells below (always 10 plus their top facets). *C*.—Like *B*, but with 4 septa removed.

When the epidermal and subepidermal cells have the same area in cross-section and occur in rectangular form, their mutual relations may be shown in the diagram figure 2, *A*, essentially like figure 1, *A*. Its heavily outlined epidermal cells are all 11-hedral, and its lightly outlined subepidermal are all 14-hedral. The removal of septum *aa*, or any one like it, makes a subepidermal cell of twice the area of the epidermal, whereupon two of the latter each lose a basal facet and become 10-hedral. The removal of another septum (*dd*) may produce two more 10-hedra (Fig. 2, *B*), or, if septum *cc* is the second to go, one 9-hedron and two 10-hedra with the same combined loss in epidermal facets. A general reduction of the epidermal cells to half the cross-sectional area of the subepidermal makes the average epidermal cell 10-hedral, and the subepidermal 16-hedral. The latter then have 4 basal facets (when the cells beneath them have the same dimensions), 6 lateral facets, and 6 toward the epidermis (Fig. 2, *B*).

The removal of septa *aa* and *bb* in figure 2, *A*, forms an underlying cell of three times the area of the related epidermal cells, four of which have become 10-hedra; or, as the process extends by taking out *cc* and *dd*, the condition in figure 2, *C*, results, essentially like that observed in *Tradescantia*. The epidermal cells one-third the size of those beneath, and presenting two 10-hedra for one 9-hedron, have an average of $9\ ^2/_3$ facets. The subepidermal cells, 8-faceted toward the epidermis, acquire an average of 18 facets.

A remarkable feature of this transformation, effected in cells by the introduction of epidermal septa and not by the removal of subepidermal, is its continued accordance with the simple formula established when the epidermal and subepidermal cells have the same average cross-sectional areas (as in Figs. 1, *A* and 2, *A*). Then, unless a space is completely covered or enclosed, if $np$ is the number of peripheral or epidermal cells, and $nc$ the number of central or subepidermal cells, their total number of facets, $F$ equals $11np$ plus $14nc$. In figure 2, *C*, for example, the pattern reduces to a multiple of 3 peripheral for 1 central, having 29 + 18 or 47 facets, whereas by the formula the same total is found as 33 + 14.

If the situation in the diagram figure 2, *C*, is considered reversed, with the lightly outlined cells as epidermal and the smaller dark ones as internal and subtended by a layer of their own size, for one epidermal with 15 facets there would be two subepidermal with 13 and one with 12 facets, 53 facets altogether. By the formula there should be $11 + (3 \times 14) = 53$.

The cucumber, however, with the same relative cell areas as *Tradescantia* which complies with the formula, affords a definite deviation. Instead of one 9-hedron for two 10-hedra as in figure 2, *C*, the cucumber, as interpreted in figure 1, *B*, would have one 8-hedron for two 10-hedra;, and in place of an average subepidermal 18-hedron it would substitute a 17-hedron. These differences are associated with the presence of enclosed cells with 1-faceted bases, not found in the patterns of figure 2. In vertical sections of the epidermis this difference is strikingly evident, since with cells showing the ordinary 2-sided bases, the cucumber has 36% of them 1-sided and 2% 3-sided, whereas in *Tradescantia* 1% are 1-sided and 24% 3-sided. These come and go with growth and cell division in a shifting pattern.[5]

The purely diagrammatic figure 3, in which the epidermal cells of one-third the sectional area of the subepidermal are shown in the rectangular form of *Tradescantia* while presenting the deviation in the cucumber, demonstrates the effect of the single basal facet. Throughout the pattern for one 8-hedral epidermal cell with a 1-faceted base, there are two 9-hedral, two 10-hedral and one 11-hedral—6 epidermal cells with 57 facets, over 2 subepidermal with 35 facets—92 altogether, where the formula calls for $66 + 28 = 94$. There is a loss of two facets, one in the epidermis and the other, its duplicate, in the subepidermis, for every cell with a single-faceted base.

This conclusion rests only on the several instances studied. In the superimposed networks of regular hexagons already cited, the smaller having areas approximately one-third of the larger, 49 "epidermal" bases are over 16 "subepidermal" tops. Whereas the corresponding cells by the formula $11np + 14nc$ should have 763 facets, the geometrical pattern, which has 9 single-faceted bases, has 745 facets, a deduction of 2 for each of them. By making 7 apothems of the small hexagons equal to 5 of the large, the small ones are about half the size (51%) of the others. When superimposed, 49 epidermal with 4 single-faceted bases are over 25 subepidermal. By the formula there should be 889 facets, and there are 881.

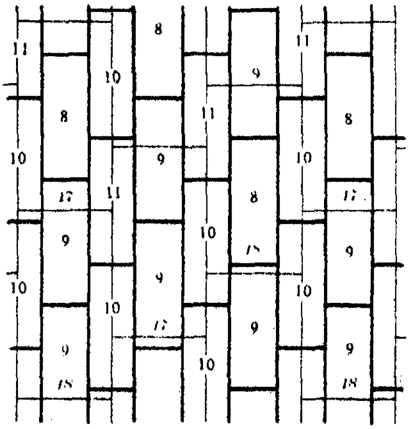


FIGURE 3

Layer of rectangular hexagons over a layer of others three times as large. The facet numbers are as in figure 2, *B*.

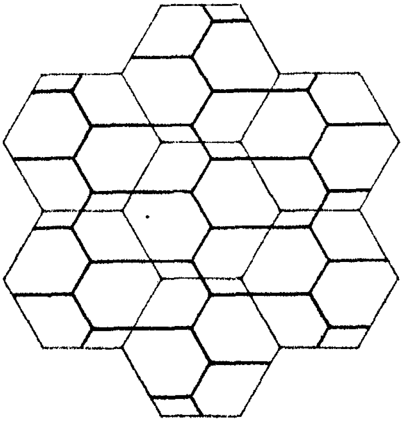


FIGURE 4

Diagram of 10-hedral epidermal cells over 16-hedral subepidermal cells of twice their area.

The extent to which cells with single-faceted bases occur in the vegetable tissues studied statistically has not been determined. Obviously impossible with cells of the same area as those beneath, and difficult to produce in those of nearly the same size, it can occur with great frequency when the areas are as 1 to 3. In a layer of small cells between two layers of large ones, arranged as in figure 1, *B*, the average facets of the small internal cells could be reduced from 14 to $10\,^2/_3$, and there are many possible intermediate reductions. Perhaps this range includes "the internal cells of the young midrib" of *Utricularia*, 100 of which average 12.34 facets per cell as studied by Miss Flint.[4]

A nearly perfect example of the epidermal geometry has been found by Matzke[6] near the stem tip of *Elodea*, where the "subepidermal cells were somewhat larger than those of the epidermis." There he found that 200 epidermal cells had an average of 10.005 facets, and that their most frequent shape, presented by 34 of them, was the 4-4-2 pattern (4 quadrilateral facets, 4 pentagonal and 2 hexagonal). In our diagram (Fig. 4) the underlying cells are twice the area of the epidermal, and all of the latter are 10-hedra of the 4-4-2 pattern. The free surfaces of the epidermal cells have the shape that Goebel drew in the little seaweed *Giraudia*,[7] which so pleased Thompson that he modified the picture and placed it in *Growth and Form* with a confusing misspelling as the "Epidermis of *Girardia*."[8] Figure 4, which is a variety of figure 2, *B*, is not presented as a picture of *Elodea*, or exactly of any other plant, but as a diagram which may help to put in order the observed confusion. The subepidermal hexagonal mosaic which presents 2 vertices (or an average of 2) to be covered by each epidermal cell of the same size, giving it a 4-faceted base, presents a single vertex yielding a 3-faceted base for epidermal cells of half the size and double the number of the subepidermal. Additional epidermal cells, finding no vertex available, may form either 2-faceted bases over an edge (*Tradescantia*), or single-faceted bases over the surface away from the corners and edges (cucumber).

*Summary.*—Because of the geometric control of cell shapes in plants, the previously reported average of 9 $^{1}/_{3}$ facets per epidermal cell resting on subepidermal cells of three times their area, in the cucumber, and 9 $^{2}/_{3}$ facets per epidermal cell resting on subepidermal cells of three times their area, in *Tradescantia*, appeared as a discrepancy in need of explanation. The 9 $^{2}/_{3}$ average, which accords with the geometrical formula whereby epidermal cells primarily average 11 facets and internal cells 14 facets, has been verified. The epidermal cells of the cucumber, revised as having an average of 9 $^{1}/_{2}$ facets, remain a deviation which depends on the occurrence of epidermal cells with single-faceted bases. For epidermal cells one-third the area of the cells beneath, the presence of those with single-faceted bases in the proportion of 1 in 6 would account for a reduction in the average epidermal facets from 9 $^{2}/_{3}$ to 9 $^{1}/_{2}$.

[1] Matzke, E. B., these PROCEEDINGS, **31**, 281–289 (1945).

[2] Lewis, F. T., *Proc. Amer. Acad. Arts and Sci.*, **77**, 147–186 (1949) [p. 176].

[3] Lewis, F. T., *Anat. Rec.*, **38**, 341–376 (1928).

[4] Flint, T. J., *Am. J. Bot.*, **36**, 397–404 (1949).

[5] Lewis, F. T., *Anat. Rec.*, **47**, 59–99 (1930).

[6] Matzke, E. B., *Am. J. Bot.*, **35**, 323–332 (1948).

[7] Goebel, K., *Bot. Zeit.*, **36**, 177–184; 193–201 (1878).

[8] Thompson, D'Arcy W., *On Growth and Form*, Cambridge, 1917, p. 321; new ed., 1942, p. 508.

# *THE EFFECT OF NITROGEN MUSTARD PRETREATMENT ON THE ULTRA-VIOLET-INDUCED MORPHOLOGICAL AND BIOCHEMICAL MUTATION RATE*[1]

BY C. P. SWANSON, W. D. McELROY AND HELMA MILLER[2]

JOHNS HOPKINS UNIVERSITY

Communicated by B. H. Willier, July 22, 1949

A wide variety of experimental studies have indicated that the stability of genes to mutagenic radiations resides, to a large extent, in the inherent nature of the individual genes, and is independent of extrinsic factors to which the genes can be experimentally exposed. While it is known that the same genes may respond differentially to x-rays when in different tissues, e.g., spermatogonia as opposed to mature spermatozoa,[3] the rather constant rate of mutability of individual genes, and even of alleles of the same gene, is somewhat surprising since a rather diverse group of supplementary treatments have been shown to influence the rate of detectable x-ray-induced chromosomal rearrangements, and since it also is known that for this type of radiation there exists a close, if unresolved, relationship between rearrangements and mutations.[4] Even when the x-ray-induced frequency of mutations can be increased by non-mutagenic supplementary treatments such as low temperatures,[5] infra-red[6] and by high versus low intensities,[7] or decreased by exposure *in vacuo*,[8] the conclusion that these data reveal a dependence of genic stability upon environmental factors is open to suspicion. Such induced mutations and, in particular, the recessive lethals as found in Drosophila, do not form a homogeneous group, some being associated with chromosome breakage, others ostensibly not, and some being reversible, others not.

The mutations induced by ultra-violet radiation are less suspect of being associated with chromosomal rearrangements, ultra-violet being incapable of inducing the variety of intra- and inter-chromosomal changes characteristically produced by x-radiation.[9, 10] Therefore, when it was found[11] that a 30-minute pretreatment of the spores of *Aspergillus terreus* with a non-mutagenic aqueous concentration (0.1%) of nitrogen mustard greatly increased the mutability of the genes to ultra-violet (2537 *A*), it appeared that the technique offered an opportunity to investigate further the possibility of materially altering the stability of the genes to subsequent exposure to ultra-violet. In the above experiment, only morphological mutations were considered, and since *A. terreus* possesses no known sexual stage, the genetic analysis could be carried no further. Consequently, some doubt remained concerning the classification of morphological mutation as true point mutations, and a more refined approach to the question of altered mutability was needed. Additional experiments were therefore carried out on a microconidial strain of *Neurospora crassa*.[12] The frequency of both morphological and biochemical mutations was determined, but rather than test the morphological mutations for their transmissibility and possible associated sterility by means of the ascospore-isolation technique, the biochemical mutations because of their homogeneity were considered to be as close an approximation to true point mutations as was possible. Most of them can be readily mapped,[13] and many of them are reversible in nature.[14] There is also less likelihood that terminal deletions, which are the type of chromosomal change predominantly induced by ultra-violet, would lead to viable biochemical mutations without associated morphological change in a haploid organism such as Neurospora.

The technique of treating the spores with ultra-violet has been described.

TABLE 1

FREQUENCY OF ULTRA-VIOLET-INDUCED MORPHOLOGICAL MUTATIONS IN *Neurospora crassa* WITH AND WITHOUT PRETREATMENT WITH NITROGEN MUSTARD

| TREATMENT, IN MINUTES | NO. ISOLATES | NO. MUTANTS | MUTATION, % | P VALUE |
|---|---|---|---|---|
| Control | 1018 | 4 | 0.39 | 0.15 |
| Mustard (0.1%) | 1308 | 12 | 0.92 | |
| UV—2* | 588 | 24 | 4.1 | <0.0001 |
| M + UV—2 | 541 | 72 | 13.3 | |
| UV—5 | 1982 | 140 | 7.1 | <0.0001 |
| M + UV—5 | 1858 | 208 | 11.2 | |
| UV—7½ | 623 | 118 | 18.9 | 0.002 |
| M + UV—7½ | 725 | 202 | 27.9 | |
| UV—10 | 1698 | 290 | 17.1 | <0.0001 |
| M + UV—10 | 1358 | 358 | 26.2 | |
| UV—12½ | 2190 | 340 | 15.5 | |
| M + UV—12½ | 1640 | 254 | 15.8 | |

* UV given at a rate of approx. 130 ergs/cm.$^2$/sec.

Eight-day old cultures were used throughout the experiment. The mustard employed was bis-beta-chloroethylmethylamine(HCl), and the 30-minute pretreatment consisted simply of suspending the microconidia in a 0.1% aqueous concentration. The spores were washed before radiation, irradiated and then plated onto a complete medium to which 3% *l*-sorbose had been added.[15] The isolates were transferred to a complete medium, scored for morphological mutations and then transferred to a liquid minimal medium for the detection of biochemical mutations. Only those isolates which exhibited the growth characteristics of the parental strain, but which did not grow on a minimal medium, were classified as possessing a biochemical mutation. Most of these could be shown to involve single amino acid or vitamin deficiencies, although a few grew only when a complex substrate was provided. Not all of the isolates were tested for biochemical mutations (compare the number of isolates in tables 1 and 2), and some of the morphological mutants were untestable biochemically because of their nonsporulating nature and thin mycelial growth.

Table 1 summarizes the data dealing with the frequency of morphological mutations. These were recognized by their effect on the growth characteristics of the colonies on a complete medium. Very few color mutants were observed because of the albinistic nature of the strain. There was no increase in the frequency of morphological mutations when transfer was made to the minimal medium, indicating that this type of mutant was not dependent upon any simple biochemical requirement.

Unlike *A. terreus*, this microconidial strain was sufficiently sensitive to the 0.1% concentration of nitrogen mustard to yield a percentage of mutations more than twice that of the untreated controls, an increase, however, which was not significant at the 1% level. Except for the $12^{1}/_{2}$-minute series, the mustard-ultra-violet treatments showed a highly significant increase in frequency of mutations over that obtained from the ultra-violet controls. These data are in substantial agreement with similar data obtained from *A. terreus*,[11] where the mustard pretreatments gave high increases at the low ultra-violet dosages, raised the peak of the mutation curve and caused a more abrupt decline in frequency at the higher dosages. The microconidia, however, were so much more sensitive to ultra-violet than were the spores of *A. terreus* that it was not possible to obtain large numbers of isolates at the higher dosage levels, with the consequence that the rate of decline in frequency for the two series could not be compared.

TABLE 2

Frequency of Ultra-Violet-Induced Biochemical Mutations in *Neurospora crassa* with and without Pretreatment with Nitrogen Mustard

| Treatment, in Minutes | No. Isolates | No. Mutants | Mutations, % |
|---|---|---|---|
| Control | 750 | 0 | 0.00 |
| Mustard (0.1%) | 1003 | 0 | 0.00 |
| UV—2 | 587 | 2 | 0.34 |
| M + UV—2 | 531 | 3 | 0.56 |
| UV—5 | 902 | 4 | 0.44 |
| M + UV—5 | 825 | 6 | 0.73 |
| UV—$7^{1}/_{2}$ | 603 | 6 | 1.00 |
| M + UV—$7^{1}/_{2}$ | 701 | 12 | 1.70 |
| UV—10 | 1698 | 17 | 1.01 |
| M + UV—10 | 1358 | 23 | 1.69 |
| UV—$12^{1}/_{2}$ | 1406 | 14 | 1.00 |
| M + UV—$12^{1}/_{2}$ | 724 | 15 | 2.07 |

$X^2 = 20.1913$ $N = 4$ $P < 0.01$

Table 2 summarizes the data dealing with the frequency of biochemical mutations. Since no comparable data of this sort have been previously published, figure 1 is included for graphic purposes. Neither the untreated nor the mustard control series yielded any mutations of this nature. A consistently higher frequency was obtained from all of the mustard-ultra-violet treatments as compared with their respective ultra-violet controls, but unlike the downward trend found in the rate of morphological mutations no fall in frequency of biochemical mutations was evident. The mustard-ultra-violet rate continued to rise throughout the range of dosages employed, while the ultra-violet controls reached a peak at the $7^{1}/_{2}$-minute dosage, and continued at the same level through the 10- and $12^{1}/_{2}$-minute intervals. Whether an eventual decline in frequency would take place at still higher dosages could not be determined because of the very low survival rate of the microconidia.

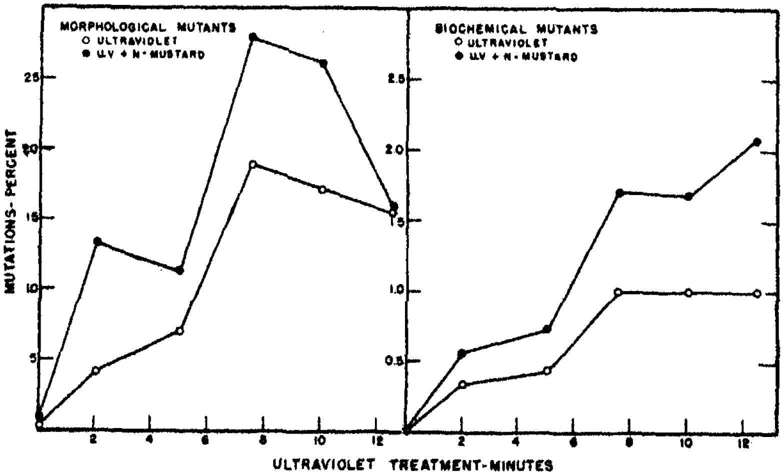


FIGURE 1

Frequency of morphological and biochemical mutations in *Neurospora crassa* as a function of ultra-violet dosage.

Of the five ultra-violet dosages recorded in table 2, only the $12^{1}/_{2}$-minute series showed a highly significant increase with the mustard pretreatment, with the 10-minute series being significant at the 5% level only. The remaining three dosages gave no individually significant increases, but the summated chi-square values for the five series, as computed by means of a five-place contingency table, yielded a P value of $< 0.01$.

It appears reasonably certain, therefore, that the frequency of both morphological and biochemical mutations can be materially altered by pretreatment with a non-mutagenic concentration of nitrogen mustard. It is not possible to state what the nature of the sensitization mechanism might be, but there can be little doubt that the frequency of mutation under these experimental conditions is not solely a function of the amount of radiant energy absorbed. Whether the altered frequency has resulted from an altered stability of the genes, or from an altered frequency of extragenic changes which have a phenotypic expression (e.g., position effects), could not be determined with absolute certainty, but the data on biochemical mutations, which represent a relatively homogeneous group, and the known action of ultra-violet on genes and chromosomes, suggest that the former alternative is more likely the correct one.

The data on the frequency of biochemical mutations as a function of dosage also have a bearing on the interpretation of the shape of the morphological mutation curve. It has been proposed that the decline in frequency of morphological mutations at the higher dosages of ultra-violet results from a selective killing action which favors the non-mutated spores, i.e., the deleterious action of the ultra-violet on the cytoplasm makes more difficult the germination and growth of the mutated spores.[6, 16, 17] On this hypothesis, the biochemical mutations, being essentially wild type when grown on a complete medium, should possess no selective disadvantage, and should continue to increase as the dosage is increased. Support for this is provided by the data in table 2 where no fall in frequency of biochemical mutations was found at dosages where the morphological frequency was declining, although a leveling off of mutation rate is found for the ultra-violet control series.

*Summary.*—Pretreatment of the microconidia of *Neurospora crassa* with a relatively non-mutagenic concentration of nitrogen mustard alters the stability of the genes in such a manner as to increase significantly the effectiveness of ultra-violet in inducing morphological and biochemical mutations. It is suggested that the increased frequency of mutations results from an increased frequency of intra-genic rather than extra-genic (chromosomal) alterations.

---

[1] Supported by grants from the National Institutes of Health, U. S. Public Health Service.

[2] Present address: Department of Anatomy, University of Chicago.

[3] Ives, P. T., and Andrews, M. B., *Genetics*, **31**, 220 (1946); Muller, H. J., *Yearbook Amer. Phil. Soc.*, **1945**, 150–153 (1946); Lamy, R., *J. Genet.*, **48**, 223–236 (1947).

[4] Lea, D. E., *Actions of Radiations on Living Cells*, p. 402, University Press, Cambridge, England (1946); Herskowitz, I. H., *Am. Nat.*, **80**, 558–592 (1946); Kaufmann, B. P., and Gay, H., these PROCEEDINGS, **33**, 366–372 (1947).

[5] Medvedev, N. H., *Acad. Sci. U.S.S.R., Compt. Rend.*, **4** (9), 283–285 (1935); King, E. D., *Genetics*, **32**, 161–164 (1947).

[6] Swanson, C. P., Hollaender, A., and Kaufmann, B. P., *Genetics*, **33**, 429–437 (1948).

[7] Sansome, E. R., Demerec, M., and Hollaender, A., *Am. J. Bot.*, **32**, 218–226 (1945).

[8] Hayden, B., and Smith, L., *Genetics*, **34**, 26–43 (1949).

[9] Stadler, L. J., *Proc. 7th Internat. Gen. Cong.*, pp. 269–276 (1939); *Cold Spring Harbor Symposia Quant. Biol.*, **9**, 168–177 (1941).

[10] Swanson, C. P., these PROCEEDINGS, **26**, 366–373 (1940).

[11] Swanson, C. P., and Goodgal, S. H., *Genetics*, **33**, 127 (1948).

[12] Our strain No. 146 was derived from a fluffy-microconidial strain No. 8743–21 (13–7) kindly supplied by Drs. Tatum and Barratt of Stanford Univ.; Barratt, R. W., Garnjobst, L., and Tatum, E. L., *Proc. 8th Internat. Gen. Congr.*, Suppl. Hereditas (1949).

[13] Beadle, G. W., Chap. VI, *Science in Progress*, 5th Series, Yale Univ. Press (1947).

[14] Ryan, F. J., *Cold Spring Harbor Symposia Quant. Biol.*, **11**, 215–227 (1946); Giles, N. H., Jr., *Genetics*, **33**, 105–106 (1948).

[15] Tatum, E. L., Barratt, R. W., and Cutter, V. M., *Am. J. Bot.*, **35**, 803 (1948).

[16] Hollaender, A., and Emmons, C. N., *Cold Spring Harbor Symposia Quant. Biol.*, **9**, 179–186 (1941).

[17] Tatum, E. L., *Ibid.*, **11**, 278–284 (1946).

---

# *THE PRINCIPLES OF TIME SERIES ANALYSES APPLIED TO OCEAN WAVE DATA**

BY H. R. SEIWELL†

WOODS HOLE OCEANOGRAPHIC INSTITUTION, WOODS HOLE, MASSACHUSETTS

Communicated by J. A. Fleming and read before the Academy, April 25, 1949

*Introduction.*—In the original report on the auto correlation analysis of ocean wave data the apparent advantages of that method over periodogram and Fourier series applications to finite amounts of data were brought out.[1] Because of practical restrictions in that only finite amounts of observational data are available for study, the fitting of wave heights as functions of time by means of trigonometric components is only in special circumstances suitable for describing the physical characteristics of the data itself. In the case of ocean wave data, *a priori* knowledge of frequency is not available, and results from this type of analysis have resulted in claims that sea surface roughness patterns are composed of large numbers of frequencies which have traveled with independent velocities from their regions of generation. In other words, mathematical representations of the data have been assumed to be physically significant, a situation which may lead to apparent misconceptions of the behavior of the phenomenon and of the dynamics causing it. In fact, the fitting of data by any of the various methods of cycle analyses, in the absence of *a priori* knowledge of the physical properties, is no more effective than the fitting of that same data, say, by orthogonal polynomials.[2]

It appears that the problem of understanding the behavior of a physical phenomenon which varies with time is directly associated with that of predicting its future course from its past behavior, and the distance into the future which the phenomenon may be accurately forecasted may be taken as an indication of the amount of dynamics associated with the phenomenon itself. Thus, in the case of sea surface wave data, while a mathematical theory may be set up from the results of a periodogram or Fourier series analysis to explain a record, it has little physical meaning unless it can be extrapolated into the future. It was from this point of view that Wiener[3] investigated the problem of prediction and defined the dynamics of a time series in terms of the possibility of prediction into the future. The auto correlation of a time series provides a means of doing this. It can be shown that the auto correlation function may be obtained from the spectrum of the series, and the spectrum may also be obtained from the auto correlation function. In this process the spectrum and the auto correlation function of a discrete sequence of observations preserve information by retaining the frequency and the amplitudes of all frequencies from zero to infinity, whereas, on the other hand, it discards information relative to phase relationships.

TABLE 1

| RECORD NO. | LOCATION | DEPTH, FT. | DATE | STARTING TIME | RECORD LENGTH, SECS. |
|---|---|---|---|---|---|
| B-713a | Bermuda | 120 | 2/21/47 | 1840 | 350 |
| B-12a | Bermuda | 120 | 10/25/46 | 1405 | 350 |
| 1-5B | Cuttyhunk | 75 | 6/14/46 | 1149 | 60 |
| 39-L | Cuttyhunk | 75 | 7/ 4/46 | 0627 | 350 |
| 53-W | Cuttyhunk | 75 | 9/15/46 | 0050 | 350 |
| 53-X | Cuttyhunk | 75 | 9/15/46 | 0650 | 350 |
| 53-Y | Cuttyhunk | 75 | 9/15/46 | 1250 | 350 |
| 56-E | Cuttyhunk | 75 | 10/ 1/46 | 1244 | 350 |

The method of auto correlation analysis is employed in the analysis of many types of scientific problems. In addition to its use as a prediction mechanism, the correlograms and the Fourier transforms of the correlograms into power spectra serve to describe the basic data by measuring its physical properties of period, amplitude and damping coefficients which, in turn, may be employed in an attempt to discover something of the physical causes. The results throw new light on the physical properties of the sea surface and on the behavior of ocean wave patterns. The method provides a means for distinguishing between mathematical and realistic physical representations of sea surface roughness patterns.

*Computation of the Auto Correlation Coefficient.*—Theoretical formulae defining the auto correlation coefficient as developed by Wiener,[3] Khinchine[4] and others are not strictly applicable to practical computations.[5] This is because wave heights are not observed as infinite functions of time (very frequently only short series of observations are obtainable), and also because the means of such finite series frequently depart from zero. In our computations it was found necessary to use a running mean rather than extracting a grand mean at the very beginning of the computation. To compute the auto correlation coefficient we compute the following statistics from $N$ equally spaced observations for $k$ equally spaced time lags.

$$T_0 = \sum_{i=1}^{N} x_i, \quad T_k = \sum_{i=1}^{N-k} x_i, \quad T_k' = \sum_{i=1}^{N-k} x_{i+k}$$

$$S_0 = \sum_{i=1}^{N} x_i^2, \quad S_k = \sum_{i=1}^{N-k} x_i x_{i+k}$$

$$S_0 = \frac{NS_0 - T_0^2}{N^2}, \quad S_k = \frac{(N-k)S_k - T_k T_k'}{(N-k)^2}$$

$$r_k = S_k/S_0$$

The correlogram is obtained by plotting the auto correlation coefficient ($r_k$) as ordinate against $k$ as abscissa. In this particular investigation the correlogram provides the basis for analysis of the data, and may be transformed into a power spectrum by a Fourier transform of the auto correlation coefficients. This spectrum is identical with that which would be obtained by periodogram analysis of an infinite series.

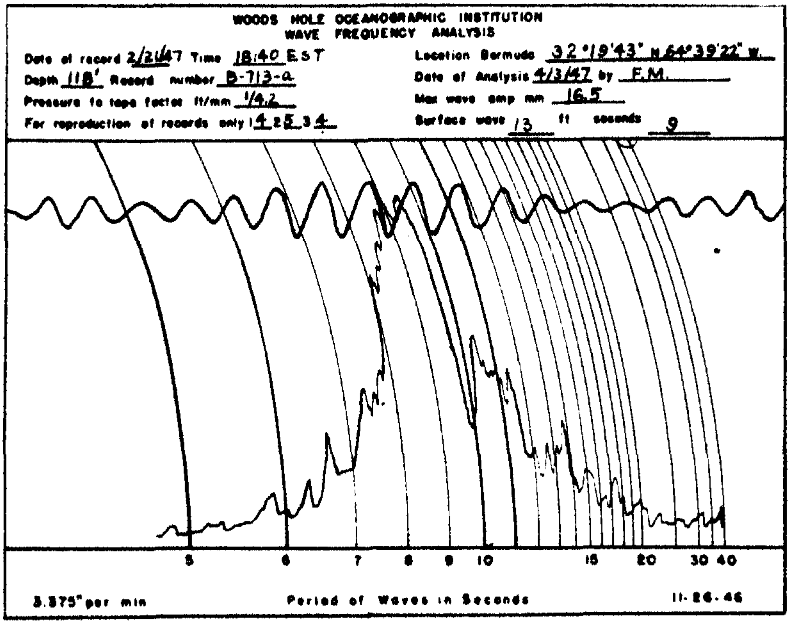
In the applications considered in this paper the auto correlation function frequently occurs as a composite of a pure sinusoidal wave and one producing a continuous spectrum. The period of the former is given by the correlogram and its amplitude computed from the original data by least squares. The residue after subtraction from the original data consists of a continuous spectrum. This provides a practical example of theorem 3 of Khinchine's theory of correlation[4] in which the auto correlation function is split into a line and a continuous spectrum.

*The Data.*—This paper presents results of auto correlation analysis of eight ocean wave records. The records are from underwater pressure recorders located at Cuttyhunk, Massachusetts (depth 75 feet), and at Bermuda (depth 120 feet) and represent pressure pulses at the sea bottom transmitted from the sea surface.[6] For the analyses, the records were scaled for pressure values at intervals of one second and were not reduced to sea surface wave heights.[7]

Bermuda Record B-713a is discussed separately. Table 1 summarizes the records analyzed.

*The Analysis of Wave Record B-713a.*—This record is discussed in more detail than the others for comparison of the results from cycle analyses with those of auto correlation analyses on finite amounts of data. Cycle analysis refers both to mechanical periodogram analysis and to the fitting of data by sums of sines and cosines. Results obtained by the former method, to which the data had already been subjected,[8] are first discussed in the light of their usual interpretation. This is then supplemented by computation of certain statistical properties of the data and then followed by fitting a series of cosine waves to the data over the critical part of the spectrum. Results obtained by the above methods are then compared with those obtained by the method of auto correlation analysis.

*a. Mechanical Periodogram Analysis:* The mechanical periodogram analysis of this 20-minute wave record (Fig. 1) shows a continuous spectral distribution with about 32 defined peaks between 8.8 and 40 seconds. The greatest development occurs in the 8- to 10-second band, and maximum energy is indicated for 8.8, 9.0 and 9.5 seconds. The customary interpretation given this type of analysis is that the basic wave pattern is made up of the frequencies indicated. The principal wave periods lie between 6 and 15 or 20 seconds, with the maximum period band of the "wave train" extending from 8 to 10 seconds. The characteristics of different wave records are usually compared by the widths and locations of the maximum band development, together with locations of the individual maximum points. Special consideration is frequently given to indications of longer periods, say above 20 seconds, which are sometimes presumed to be waves generated by oceanic storms at great distances but which have "out-run" the storm itself.



WOODS HOLE OCEANOGRAPHIC INSTITUTION
WAVE FREQUENCY ANALYSIS

Date of record 2/21/47 Time 18:40 EST
Depth 118' Record number B-713-a
Pressure to tape factor ft/mm 1/4.2
For reproduction of records only 1 4 2 5 3 4
Location Bermuda 32°19'43" N 64°39'22" W.
Date of Analysis 4/3/47 by F.M.
Max wave amp mm 16.5
Surface wave 13 ft seconds 9

FIGURE 1

Results of mechanical periodogram analysis of Wave Record B-713a. See text.

The above physical interpretation is doubtful in that periodogram analyses of finite amounts of data are not necessarily physical, but may be only mathematical representations of the data. Its acceptance as a physical representation without *a priori* reason may lead to a misconception of both the physical properties and the causative mechanism of the sea surface roughness. This question is discussed later after additional results are developed.

*b. Statistical Characteristics:* The frequency distribution of periods, compiled from measurements of the times between successive wave crests on the record, has the following properties:

Standard deviation = 1.135 seconds
Mean = 8.59 ± 0.077 second
Mode = 8.75 seconds

Thus, the mean and modal values of distances between wave crests in the original record lie in the range of maximum energy brought out by the mechanical periodogram analysis (Fig. 1).

TABLE 2

| PERIOD | RELATIVE AMPLITUDE | PHASE ANGLE, ° |
|---|---|---|
| 7.50 | 1.00 | 156.6 |
| 8.00 | 2.31 | 111.9 |
| 8.25 | 2.97 | 23.6 |
| 8.50 | 4.01 | 30.1 |
| 8.75 | 4.60 | 18.9 |
| 9.00 | 5.29 | 69.6 |
| 9.25 | 10.93 | 276.9 |
| 9.50 | 3.49 | 231.2 |
| 9.75 | 2.20 | 210.2 |
| 10.00 | 0.77 | 16.9 |

Amplitude and phase values obtained by fitting cosine waves to Wave Record B-713a.

*c. Cosine Wave Analysis:* Results obtained by independent least square fits of ten cosine waves (periods from 7.5 to 10.0 seconds) to the basic data are tabulated in Table 2. The spectrum is similar to that obtained by mechanical periodogram analysis, increasing from a minimum at 7.5 seconds to a maximum value at 9.25 seconds and then diminishing rapidly to a minimum at 10.0 seconds. This representation, obtained from a finite amount of data, is like that of the mechanical periodogram analysis—of mathematical significance only and without physical reality. The rôle of this type of analysis is a secondary one, to be applied after the period, or periods, have been established by other means.

*d. Auto Correlation Analysis:* The correlogram of the primary data of Wave Record B-713a (Fig. 2) maintains a period of 8.75 seconds and damps

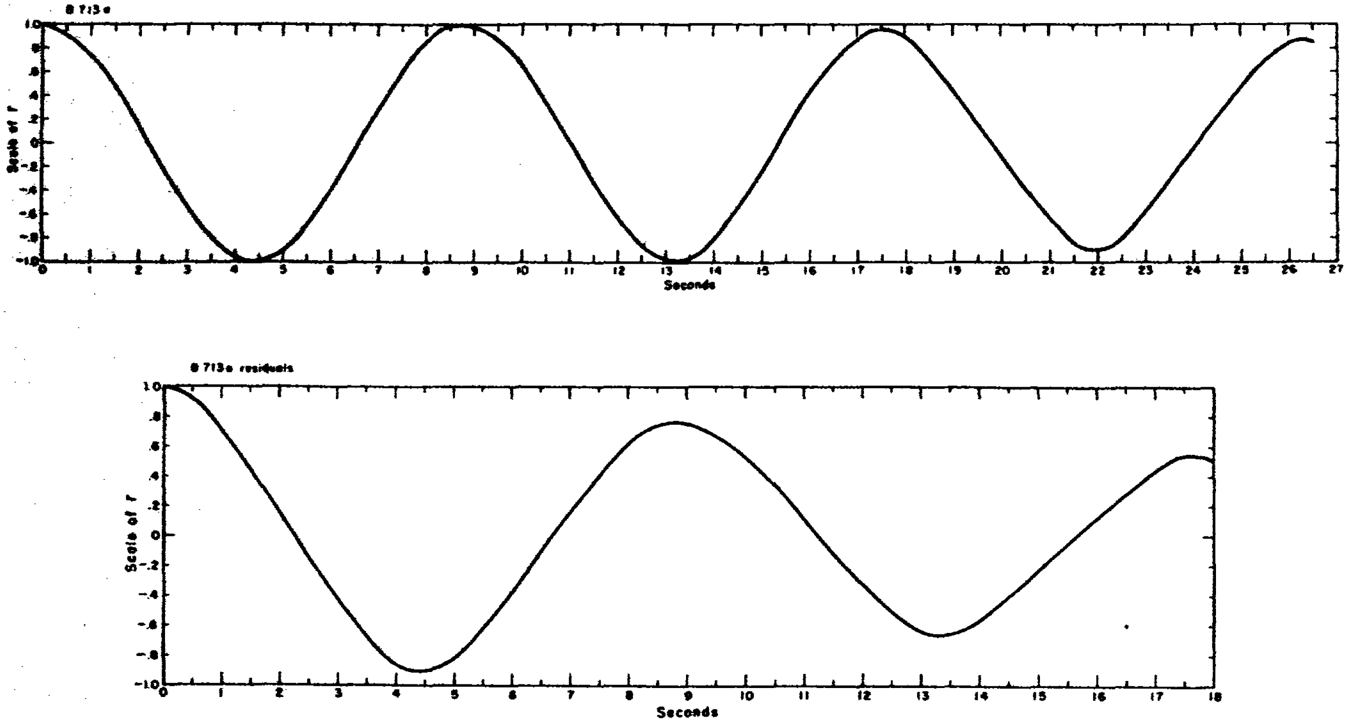


FIGURE 2

Correlograms of Wave Record B-713a. Curves have been drawn through values of $r$ computed for each second. The curves have not been smoothed.

slowly; at the end of the third cycle the amplitude is reduced to a value of $r = 0.87$, after which it continues to damp slowly until reaching a terminal value ($r_T$). The terminal amplitude is maintained by a cyclical component which proceeds undamped through the basic data. It is the fundamental wave to which energy is supplied by an external force at a rate equal to that at which it is dissipated in the system. Its amplitude, obtained from a least square fit of an 8.75-second sine wave to the primary data, is 0.720 foot.

Subtracting the cyclical component ($T = 8.75$ seconds, $A = 0.720$ foot) from the original data, the variance reduces from a value of 0.492 square foot to 0.216 square foot for the residuals. Recomputing the auto correlations of the residuals, the correlogram then damps in a fashion characteristic of an autoregressive series; at the end of the second cycle, the correlogram amplitude is $r = 0.54$ as compared to $r = 0.95$ for that of the primary data (Fig. 2).

The value of the terminal amplitude of the original correlogram is

$$r_T = \frac{\text{Variance cosine}}{\text{Total variance}} = \frac{A^2/2}{\sigma_y^2} = 0.56$$

for Wave Record B-713a. This value also represents the per cent reduction due to the cyclical component itself, and may be expressed as the ratio of the residual ($\sigma_r^2$) and original ($\sigma_y^2$) variances. Thus:

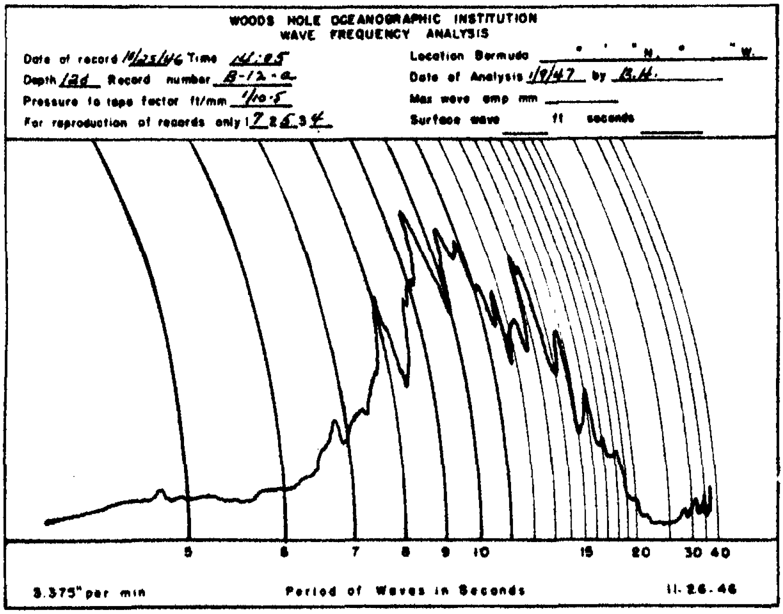
$$r_T = 1 - \frac{\sigma_r^2}{\sigma_y^2}$$

The numerical solutions of the above and other relations, for cases where the original series consists of a sine term and a random residual, provides a means of checking the computations.

The significant difference between the results of periodogram analyses and correlogram analyses of finite amounts of data is that the former are not the power spectra of the latter. The large number of peaks in the periodogram of Record B-713a (Fig. 1) are not real periods, and they provide only a mathematical representation of the data. The spectrum is similar to that obtained by the arbitrary fitting of cosine waves to the data, and is not to be interpreted as having physical reality.

On the other hand, the use of auto correlation analysis reveals that for Record B-713a the auto correlation function is broken into two parts—one having a line, and the other a continuous spectrum. The line spectrum is that of the cyclical component present in the basic data, whereas the continuous spectrum represents a wide band of frequencies, with an average frequency equal to that of the cyclical components, but which damp out rather rapidly.

The physical interpretation to be given results from the two types of analyses is equally significantly different. The auto correlation analysis suggests that the physical characteristics of the sea surface wave pattern comprise one fundamental wave with a period of 8.75 seconds and an amplitude of 0.72 foot, on which is superimposed a series of local impulses. The former, generated presumably by a meteorological situation acting at some distance from the target, is the predictable component in the data. It accounts for an estimated 56% of the total variability of the sea surface pattern (obtained from the value of the terminal amplitude of the correlogram) and lends itself to being forecast from knowledge of wind action on the sea surface. The remainder of the data represents a superimposed series of local impulses which die out rather rapidly. They would appear to be generated by local winds in the area of the target, and their relative significance to the pattern is also evaluated from the terminal amplitude of the original correlogram.



WOODS HOLE OCEANOGRAPHIC INSTITUTION
WAVE FREQUENCY ANALYSIS

Date of record 10/25/46 Time 14:25
Depth 12d Record number B-12-a
Pressure to tape factor ft/mm 1/10.5
For reproduction of records only 1 7 2 6 3 4
Location Bermuda ° ' " N. ° ' " W.
Date of Analysis 1/9/47 by B.H.
Max wave amp mm
Surface wave ft seconds

FIGURE 3

Results of mechanical periodogram analysis of Wave Record B-12a. See text.

*Auto Correlation Analyses of Additional Ocean Wave Records.*—The results of auto correlation analyses of seven additional ocean wave records from Cuttyhunk and Bermuda are tabulated in Table 3. The mechanical periodogram analyses of these records gave results similar to those for Record B-713a, each showing approximately 30 peaks. Figures 1 and 3 are representative of this analysis.

The original correlograms do not damp completely (Figs. 2 and 4),[9] terminal amplitudes range from 0.08 to 0.66 and all show well-defined periods ranging from 6.50 to 15.5 seconds. Amplitudes of the cyclical components associated with the data range from 0.10 to 0.72 foot, and their relative quantitative significance to the total sea surface roughness patterns are revealed by the terminal amplitudes ($r_T$). Five of the records show $r_T$ values exceeding 0.5, indicating more than 50% of the roughness pattern resulting from the cyclical component. For the remainder, local influences dominate the pattern, and in one of these, 53-Y, the terminal amplitude of the correlogram is so low (0.08) as to indicate the data to be almost autoregressive. In the complete absence of sea swell the original correlogram damps rapidly to zero.[10]

TABLE 3

| RECORD NUMBER | ORIGINAL DATA | | | | | RESIDUAL DATA | | |
|---|---|---|---|---|---|---|---|---|
| | $\sigma_y^2$ | $\frac{AD}{\sigma}$ | $T$ | $A$ | $r_T$ | $\sigma_r^2$ | $\frac{AD}{\sigma}$ | $A'$ |
| B-713a | 0.492 | 0.78 | 8.75 | 0.72 | 0.56 | 0.216 | 0.78 | 0.74 |
| B-12a | 0.018 | 0.80 | 9.07 | 0.10 | 0.22 | 0.014 | 0.83 | 0.09 |
| 53-W | 0.109 | 0.80 | 15.50 | 0.35 | 0.57 | 0.047 | 0.79 | 0.35 |
| 53-X | 0.185 | 0.84 | 12.25 | 0.51 | 0.66 | 0.064 | 0.77 | 0.49 |
| 53-Y | 0.070 | 0.82 | 11.00 | 0.10 | 0.08 | 0.065 | 0.80 | 0.10 |
| 56-E | 0.128 | 0.80 | 6.50 | 0.38 | 0.56 | 0.057 | 0.80 | 0.38 |
| 39-L | 0.049 | 0.82 | 8.29 | 0.15 | 0.25 | 0.035 | 0.80 | 0.15 |
| I-5B | 0.030 | 0.84 | 8.00 | 0.19 | 0.63 | 0.011 | 0.83 | 0.19 |

Results of auto correlation analyses of ocean wave records. $\sigma_y^2$ = variance original data, $AD/\sigma$ = ratio of average deviation to standard deviation, $T$ = period of correlogram, $A$ = amplitude of cyclical component by harmonic analysis, $\sigma_r^2$ = variance residual data, $r_T$ = terminal amplitude of original correlogram, $A' = \sqrt{2(\sigma_y^2 - \sigma_r^2)}$.

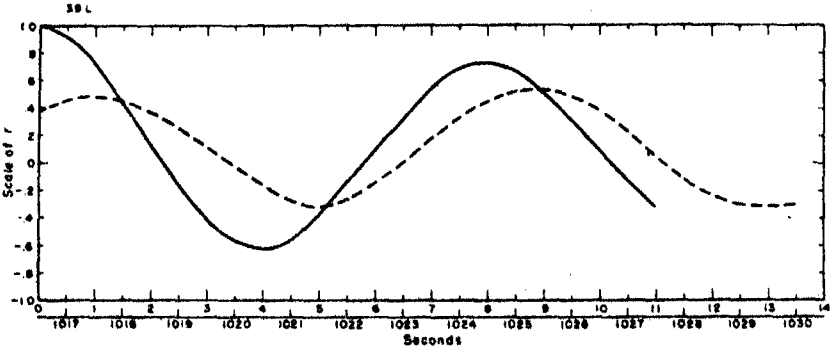


FIGURE 4

Correlogram of Wave Record 39-L, 0 to 11 seconds (solid line) and 1017 to 1030 seconds (dashed line). Curves have been drawn through values of $r$ computed for each second. The curves have not been smoothed.

The results are typical of those obtained from the analyses of more than 30 wave records from the North Atlantic, Mediterranean and the eastern American coast. After extraction of the cyclical components, the residual variances are diminished, and the correlograms of the residuals damp more rapidly in the fashion of an autoregressive series (Fig. 2).

A numerical result of apparent significance is that the ratios of the mean deviations to the standard deviations (for both original and residual data) is approximately equal to $\sqrt{2/\pi}$, a relationship which usually characterizes unimodal curves approaching symmetry. This relationship enables an approximate estimate of the mean sea surface roughness (expressed as twice the average deviation of the original data) to be made directly from the variance of the original data.

The properties of the sea surface roughness pattern, as brought out by auto correlation analysis of a finite amount of data, describe its physical characteristics. They indicate that the instantaneous sea surface pattern is not an interference pattern composed of many wave frequencies which have traveled with independent velocities from their generation areas, such as may be concluded from results of a cycle analysis of finite amounts of data, but rather that it seems to consist of a single cyclical component (sea swell) on which is superimposed random frequencies (local sea) of the same average frequency. The sea swell appears to have been generated by a dominating meteorological situation at some distance from the target and is maintained by energy supplied at a rate equal to that dissipated by the system. It is to be expected that this period changes with time and at a rate dependent on the stability of the off-shore meteorological situation causing it. On the other hand, the local sea is induced by local winds and other factors tending to disturb the sea surface and dies out fairly rapidly.

It may thus be reasoned that the ocean itself acts as a filter which, under specific conditions, passes only certain fundamental frequencies, which we call sea swell. This fundamental frequency at any time is perhaps a function of the characteristics of the ocean basin over which it has traveled, plus wind velocity, direction, fetch and other properties of an established wind-generating system. Its constancy is related to the stability of the meteorological situation. At any particular time local conditions may also produce a continuous band of frequencies with an average frequency the same as that of the sea swell. These latter die out rapidly in a manner similar to the effect of random impulses on a sluggish oscillator as the ocean basin continues to act as a filter.

However, regardless of the exact nature of the physical explanation of the mechanism maintaining the state of the sea, a quantitative separation into sea swell and local sea is fundamental to an understanding of its physical properties. This has an added practical significance for operational purposes as well as for aerial depth determination and for controlled experiments involving sea surface properties. To this end, the method of auto correlation analysis is an improvement over other procedures for the analysis of finite amounts of data in which not even the length of the periods or the recurrence intervals are known or suspected at the outset. The results in themselves provide a more realistic geophysical explanation of the phenomena and avoid the pitfalls from claims of large number of periods and cycles which may arise from procedures less adapted to the available data.

---

[*] Contribution No. 486 from the Woods Hole Oceanographic Institution and the Office of Naval Research, Contract N6onr 277/1.

[†] The methods and conclusions in this paper have partly resulted from a collaboration with Dr. George P. Wadsworth, of the Massachusetts Institute of Technology, who is preparing a theoretical account of the adaptations of generalized harmonic analysis to the specific problems of ocean waves.

[1] Seiwell, H. R., and Wadsworth, George P., "A New Development in Ocean Wave Research," *Science*, **109**, 271–274 (1949).

[2] See also: Bartels, J., "Some Aspects of Geophysical Cycles," *J. Wash. Acad. Sci.*, **26**, No. 5 (1936); Wilson, Edwin B., "The Periodogram of American Business Activity," *Quart. J. Economics*, 375–417 (May, 1934).

[3] Wiener, N., "Generalized Harmonic Analysis," *Acta Math.*, **55**, 117 (1930). "The Extrapolation, Interpolation and Smoothing of Stationary Time Series," NDRC Report, 1942. Prepublication announcements of the later report have been made by John Wiley and Sons, Inc., New York.

[4] Khinchine, A., "Korrelationstheorie der Stationaren Stochastischen Prozesse," *Math. Ann.*, **109**, 604–615 (1934). Translation by T. C. Duke, Dept. of Math., Mass. Inst. Tech., 1949.

[5] See also: Kendall, Maurice G., "The Advanced Theory of Statistics," London, 1948; Bartlet, M. S., *et al.*, "Symposium on Auto Correlation in Time Series," *Supp. J. Royal Statistical Soc.*, **VIII**, No. 1, 27–97 (1946).

[6] Seiwell, H. R., "Results of Research on Surface Waves of the Western North Atlantic," papers in *Phys. Ocean. Meteor.*, **X**, No. 4, 1–56 (1948).

[7] The ratio of amplitudes of instantaneous pressure fluctuations at the sea bottom $\Delta P_h$ to those at the sea surface $\Delta P_s$ is

$$\frac{\Delta P_h}{\Delta P_s} = 1.35 \frac{1}{\text{Cosh } kh}$$

where $h$ is depth to bottom and $k = 2\pi/\text{wave-length}$. The factor 1.35 has been experimentally determined for the Woods Hole–Bermuda area.

[8] These analyses were performed on the ocean wave analyzer at this Institution. The term mechanical periodogram analysis is used to distinguish it from other types of periodogram analyses. See reference 6.

[9] Correlograms for Record 53-X have been published in reference 1.

[10] Examples of this situation have been encountered, although they are not discussed in this paper.

# THE CLASS NUMBER OF THE CYCLOTOMIC FIELD

By N. C. Ankeny and S. Chowla

Princeton University and Institute for Advanced Study

Communicated by Marston Morse, July 21, 1949

1. Let $g$ denote any odd prime and $h = h(g)$ the class number of the cyclotomic field $R(\zeta)$, where $\zeta$ is the primitive $g$th root of unity, $R$ the rational numbers. It is known that we can write: $h = h_1 h_2$, where $h_1$ and $h_2$ (both integers) are the so-called first and second factors of the class number; in fact $h_2$ is the class number of the real field of degree 2 under $R(\zeta)$, namely the field $R(\zeta + \zeta^{-1})$.

Kummer conjectured (*J. Math.*, **16**, 473 (1851)) that

$$h_1 \sim \frac{g^{(g+3)/4}}{2^{(g-3)/2}\,\pi^{(g-1)/2}} = G. \qquad (1)$$

He also calculated $h_1$ for $g \leqq 97$ and found $h_1 = 1$ for $g \leqq 19$, $h_1 = 411322823001$ for $g = 97$. No proof of (1) has yet been published.

In this paper we prove the following result:

$$\log\left(\frac{h_1}{G}\right)/\log g \to 0 \qquad (g \to \infty). \qquad (2)$$

An interesting consequence of (2) is that there exists a $g_0$ such that $h_1(g)$ is (strictly) monotonic increasing for $g > g_0$; in fact if $g_2 > g_1 > g_0$, we have

$$h_1(g_2) > h_1(g_1).$$

Let $p$ stand for a typical prime. Define

$$c_p = +1,\ -1,\ 0 \qquad [p \equiv 1(g),\ p \equiv -1(g),\ p \not\equiv \pm 1(g)]$$

It is well known that the series

$$\Sigma c_p \cdot p^{-1}$$

is convergent. We can show that *if Kummer's conjecture is true, we must have:*

$$\Sigma c_p \cdot p^{-1} = O(g^{-1}) \qquad [g \to \infty]. \qquad (3)$$

It is hard to say whether (3) is even probably true. We cannot prove (3) even on the assumption of the "extended Riemann hypothesis" (or e. R. h.).

The method used to prove the above results can also be applied to prove the following results:

Let $\theta_1$ and $\theta_2$ denote any fixed constants such that

$$1/2 < \theta_1 < s < \theta_2 < 1.$$

If the e. R. h. is true, there exists for every given $\epsilon > 0$, a non-principal character $X(n) (\text{mod } g)$ such that

$$|L(s)| = |\sum_{1}^{\infty} X(n) \cdot n^{-s}| < 1 + \epsilon \tag{4}$$

for all primes $g > g_0(\epsilon, \theta_1, \theta_2)$.

This result with the larger constant $\sqrt{\zeta(2s)} + \epsilon$ on the right hand side of the inequality (4) (here, $\zeta(s)$ denotes Riemann's Zeta function; hence $\zeta(2s) > 1$ for $s > {}^1/_2$) is implicit in some recent work of Atle Selberg, but it (with the larger constant) is proved by him without any hypothesis (*Skrifter Norske Videnskaps-Akad. Oslo*, Nr. 3 (1946)).

2. We give only a sketch of the methods used to prove the above results. Complete proofs will be published elsewhere.

*Notation:* We write

$$a \equiv b(c) \text{ for } a \equiv b \ (\text{mod } c)$$

Let $X_t$ denote the ${}^1/_2(g - 1)$ characters (mod $g$), such that $X_t(-1) = -1$. $\pi(x; k, l)$ denotes the number of primes $\equiv l(k)$ and not exceeding $x$; $\phi(k)$ is the number of positive integers not exceeding $k$ and prime to $k$.

$X_t(n)$ is a non-principal character (mod $g$), hence

$$L_t(s) = \sum_{1}^{\infty} X_t(n) \cdot n^{-s} \qquad [R(s) > 0]$$

It now easily follows that for $s > 1$ we have

$$\pi_{t} L_t(s) = \exp \left\{ g' \sum_{p} \sum_{m \geq 1} \frac{c_{p,\,m}}{m p^{ms}} \right\} \tag{5}$$

where $g' = {}^1/_2(g - 1)$ and

$$c_{p,\,m} = +1, -1, 0 [p^m \equiv 1(g), p^m \equiv -1(g), p^m \not\equiv \pm 1(g)].$$

From (5) and

$$\frac{h_1}{G} = \pi_{t} L_t(1) \tag{6}$$

we deduce that

$$\frac{h_1}{G} = \exp \left\{ g' \sum \frac{c_p}{p} + g' \sum_{p} \sum_{m > 2} \frac{c_{p,\,m}}{m p^m} \right\} \tag{7}$$

Write

$$C = \sum c_p \cdot p^{-1} = C_1 + C_2 + C_3 \tag{8}$$

where for any given $\epsilon > 0$:

$C_1$ is the contribution to $C$ of the terms with $2g - 1 \leqq p < g^{1+\epsilon}$

$C_2$ is the contribution to $C$ of the terms with

$$g^{1+\epsilon} < p < \exp(g^{2\theta});$$

$C_3$, the rest.

It is trivial that ($B$ is an absolute positive constant)

$$|C_1| \leqq B \frac{\epsilon}{g} \log g. \tag{9}$$

Now write

$$C(x) = \sum_{p \leqq x} C_p$$

Then the sum $C_2$ is essentially

$$\sum \frac{C(n)}{n(n+1)} \tag{10}$$

where the summation is for all integers $n$ lying between

$$g^{1+\epsilon} \text{ and } \exp(g^{2\theta}).$$

To deal with the sum (10) we use the following result:

$$\pi(x;\ k, l) = O\left\{\frac{x}{\phi(k)\log x}\right\} \qquad [x \geqq k^{1+\epsilon}]. \tag{11}$$

This result follows from the work of Viggo Brun, but was first implicitly stated by Titchmarsh (*Rendiconti Circolo Mat. Palermo*, **54**, 414–429 (1930)). The constant implied in the O symbol is independent of $x$ and $k$ but may depend on $\epsilon$.

From (11) one easily deduces that

$$|C_2| \leqq \frac{A\theta K(\epsilon)\log g}{g} \tag{12}$$

where $A$ is an absolute positive constant.

Finally to estimate $C_3$ we use the following theorem (*Walfisz, Math. Ztschr.*, **40**, 598–599 (1936)): If $(k, l) = 1$, then

$$\pi(x;\ k, l) - \frac{1}{\phi(k)}\int_2^x \frac{du}{\log u} = O\left\{\frac{x}{\log^m x}\right\} + O\left\{\frac{x^{1 - ck^{-\theta}}}{\phi(k)\log x}\right\}$$

where the constant implied in the O symbol is independent of $x$ and $k$, but may depend on $m$ (arbitrarily large) and $\theta$ (arbitrarily small). By the application of this result we can easily obtain:

$$C_3 = O[g^{(1-m)2\theta} + g^{-1-2\theta}] \tag{13}$$

We first fix $\epsilon$ (small), then $\theta$ sufficiently small, and finally $m$ sufficiently large, to obtain from (8), (9), (12) and (13):

$$\left(\frac{g}{\log g}\right) \Sigma c_p \cdot p^{-1} \rightarrow 0 \ (g \rightarrow \infty) \tag{14}$$

We next indicate the proof of

$$C_4 = \sum_{p} \sum_{m \geq 2} \frac{c_{p,m}}{mp^m} = O(g^{-1}) \tag{15}$$

(14) shows that (2) follows from (15). In fact,

$$|C_4| \leq \sum_{p} \sum_{m \geq 2} \frac{1}{mp^m} \qquad [p^m \equiv \pm 1(g)]$$

The part of this sum with $p > g$ is shown to be $0(g^{-1})$ by a very short argument.

Since $X^m \equiv 1(g)$ has at most $m$ incongruent solutions, we derive that

$$C_4 = O\left\{ \sum_{m=2}^{\infty} \frac{1}{m}\left[ \sum_{a=A(m)}^{B(m)} \left(\frac{1}{ag+1} + \frac{1}{ag-1}\right)\right]\right.$$

where

$$A(m) = {}^1/_2(m^2 - m) + 1, \qquad B(m) = {}^1/_2(m^2 + m)$$

And this indicates the proof of (15).

To prove (4) we start from the result

$$C(X) = O(\sqrt{x} \log x)$$

which is true on the assumption of the e. R. h. The rest of the arguments are similar to those used above.

Also, on the assumption of the e. R. h. we can improve (2) to

$$\log\left(\frac{h_1}{G}\right) = O\{(\log g)^{1/2} \log \log g\}$$

# SOME PROPERTIES OF A SFIELD

BY LOO-KENG HUA

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS

Communicated by H. S. Vandiver, July 12, 1949

All the results of this paper are initiated from the almost trivial identity that if $ab \neq ba$, then

$$a = (b^{-1} - (a - 1)^{-1}b^{-1}(a - 1))(a^{-1}b^{-1}a - (a - 1)^{-1}b^{-1}(a - 1))^{-1} \tag{1}$$

Among these results, there are two interesting ones which are the perfect form of two theorems due to H. Cartan[1] and J. Dieudonné,[2] respectively.

THEOREM 1. *Every sfield is generated by a non-central element and its conjugates.*

Let $K$ be the sfield and $L$ be the conjugate set. The identity (1) asserts that if there is an element $b$ of $L$ such that $ab \neq ba$, then $a$ belongs to the sfield $K$, generated by $L$. That is, $K - K_1$ and $K_1$ are commutative elementwise.

Suppose our theorem is false; we have an element $a$ of $K - K_1$ and two elements $b$ and $b'$ of $K_1$ such that $bb' \neq b'b$. Since $a$ and $ab$ belong to $K - K_1$, we have $(ab)b' = b'(ab) = ab'b$, which contradicts $bb' = b'b$. The theorem follows.

An immediate consequence of Theorem 1 is the following interesting result:

THEOREM 2. *Every proper normal subsfield of a sfield is contained in the center.*

H. Cartan proved the theorem under the assumption that the rank of the sfield over its center is finite. His proof is far more complicated than the present one.

More precise results can be obtained by specializing the subgroup.

THEOREM 3. *Every sfield, which is not a field, has no proper subsfield containing all its commutators.*

It is enough to prove the existence of a commutator which does not belong to the center. Suppose $ab \neq ba$. From a modification of the identity (1):

$$a = (1 - (a - 1)^{-1}b^{-1}(a - 1)b)(a^{-1}b^{-1}ab - (a - 1)^{-1}b^{-1}(a - 1)b)^{-1}, \tag{2}$$

we deduce that at least one of $a^{-1}b^{-1}ab$ and $(a - 1)^{-1}b^{-1}(a - 1)b$ does belong to the center.

As corollaries of Theorem 3, we have

THEOREM 4. *If the center of a sfield contains all the commutators of the* *sfield, then the sfield is commutative. An element of a sfield commutative with all the commutators of the sfield belongs to the center.*

Moreover, as a simple consequence of

$$a^{-1}c^{-1}ac = a^{-2}(ac^{-1})^2c^2 \tag{3}$$

and Theorem 3, we have

THEOREM 5. *Every sfield, which is not a field, has no proper subsfield containing all its square elements.*

THEOREM 6. *If the center of a sfield contains all the square elements of the sfield, then the sfield is commutative. An element of a sfield commutative with all the square elements of the sfield belongs to the center.*

By means of an identity in the theory of finite differences:

$$\Delta^{k-1}x^k = k!(x + {}^1/_2(k - 1)),$$

where $\Delta f(x) = f(x + 1) - f(x)$ and $\Delta^i f(x) = \Delta(\Delta^{i-1}f(x))$, we have

THEOREM 7. *Every sfield of characteristic $\geq k$, which is not commutative, has no proper subsfield containing all its $k$-th power elements.*

Let us denote $a^{-1}b^{-1}ab$ by $(a, b)$. An $r$-commutator $(r \geq 2)$ is defined inductively by the relation

$$(a_1, \ldots, a_r) = (a_1, (a_2, \ldots, a_r)). \tag{4}$$

Let $C_r$ be the set of all $r$-commutators of the sfield.

THEOREM 8. *Let $r \geq 2$. Every sfield, which is not commutative, has no proper subsfield containing all its $r$-commutators. An element of a sfield which permutes with all the $r$-commutators of the sfield belongs to the center.*

Let $A_r$ and $B_r$ denote the two propositions of our theorem. $B_r$ follows from $A_r$ immediately. Theorems 3 and 5 assert that both $A_2$ and $B_2$ are true. We shall prove the theorem by induction, we assume that $A_{r-1}$ and $B_{r-1}$ are both true. Let $a$ be an element of the sfield, which does not belong to the center. By $B_{r-1}$, we have an element belonging to $C_{r-1}$ such that a $b \neq ba$. The identity (2) asserts that $a$ belongs to the sfield generated by $C_r$. Therefore $A_r$ is true. The theorem is proved.

The theorem asserts that the lower central series of the multiplicative group of a sfield can never be ended at the identity. It is an almost trivial fact that the upper central series can never go up. It is a comparatively difficult question about the derived series. Elsewhere I shall prove

THEOREM 9. *Let $r \geq 1$. Every sfield is generated by all the elements of the $r$-th derived group.*

An application of Theorem 1 is to establish the simplicity of the projective special linear group $PSL_n(K)$[3] of dimension $n$ over the sfield $K$.

THEOREM 10. *The projective special linear group $PSL_n(K)$ is simple except when $n = 2$ and $K$ has 2 or 3 elements.*

In case $K$ is a field of characteristic $\neq 2$ or $K$ is a complete field, the theorem was proved by Burnside-Jordan-Dickson (see Dickson,[4] p. 85 and Van der Waerden,[5] p. 7). For the incomplete field, it was proved by Iwasawa.[6] For a sfield, Dieudonné[2] proved the theorem for $n > 2$, and for $n = 2$[7] except in the case when the center of $K$ has 2, 3 or 5 elements. Since both exceptional cases in Theorem 10 are not simple, it is the final answer to the problem about the simplicity of $PSL_n(K)$.

Owing to the previous known results, we assume, from now on, that $n = 2$ and the sfield is not commutative. It is the same thing to prove that

THEOREM 11. *Every normal subgroup of* $SL_2(K)$ *is contained in the center, except when* $K$ *has 2 or 3 elements.*

Let $A$ be an element of $SL_2(K)$ which does not belong to the center. We are going to prove that $A$ and its conjugates generate $SL_2(K)$. We shall prove that there is a matrix $Q$ belonging to $SL_2(K)$ such that $QAQ^{-1}$ has a non-zero element at (1, 2) position. In fact, let $A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$. If $\beta \neq 0$, $Q = I$ satisfies our requirement, and if $\gamma \neq 0$, then $Q = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ does. If $\beta = \gamma = 0$, we have an element $x$ such that $\alpha x \neq x\delta$, since $A$ does not belong to the center. Then

$$\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \alpha & 0 \\ 0 & \delta \end{pmatrix}\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} * & -\alpha x + x\delta \\ * & * \end{pmatrix}.$$

We may assume that the subgroup generated by $A$ and its conjugates contains an element with $\beta \neq 0$. Since

$$\begin{pmatrix} 1 & 0 \\ \beta^{-1}\alpha & 1 \end{pmatrix}\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}\begin{pmatrix} 1 & 0 \\ \beta^{-1}\alpha & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & * \\ * & * \end{pmatrix},$$

we have assumed that the group contains an element of the form $A = \begin{pmatrix} 0 & \beta \\ \gamma & \delta \end{pmatrix}$. We take

$$B = \begin{pmatrix} -\gamma^{-1}\delta\beta^{-1} & \gamma^{-1} \\ -\gamma\kappa & 0 \end{pmatrix},$$

where $\kappa$ belongs to the commutator subgroup of $K$, so that $B$ belongs to $SL_2(K)$. We have

$$A(BA^{-1}B^{-1}) = AB(BA)^{-1} = \begin{pmatrix} -\beta\gamma\kappa & 0 \\ * & -(\gamma\kappa\beta)^{-1} \end{pmatrix}.$$

Now we shall prove that we can choose $\kappa$ such that $-\beta\gamma\kappa$ does not belong to the center. For otherwise $(-\beta\gamma\kappa_1)^{-1}(-\beta\gamma\kappa_2) = \kappa_1^{-1}\kappa_2$ belongs to the center for any two commutators $\kappa_1$ and $\kappa_2$, that is, the commutators of $\kappa$ belong to the center of $K$. By Theorem 4, $K$ is a field which contradicts our supposition. Therefore the subgroup contains an element of the form $C = ABA^{-1}B^{-1} = \begin{pmatrix} a & 0 \\ c & d \end{pmatrix}$, where $a$ does not belong to the center. We have an element $\lambda$ of $K$ such that $d\lambda \neq \lambda a$, then the subgroup contains

$$\begin{pmatrix} a & 0 \\ c & d \end{pmatrix}\begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix}\begin{pmatrix} a & 0 \\ c & d \end{pmatrix}^{-1}\begin{pmatrix} 1 & 0 \\ -\lambda & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ d\lambda a^{-1} - \lambda & 1 \end{pmatrix}.$$

We may suppose that our subgroup contains $\begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}$, $c \neq 0$.

We are going to prove that the subgroup contains $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$. In fact, if $c = \pm 1$, our assertion is trivial. Let us suppose $c \neq 0, \neq \pm 1$. The field $K_1$, obtained from the center of $K$ by adjunction of $c$, contains more than 3 elements. $SL_2(K_1)$ is generated by any one element, not in the center, and its conjugates; in particular, $\begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}$ and its conjugates generate $SL_2(K_1)$. Thus $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ is in the subgroup under consideration. Consequently the subgroup contains

$$\begin{pmatrix} \lambda^{-1} & 0 \\ 0 & \lambda \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} \lambda^{-1} & 0 \\ 0 & \lambda \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ \lambda^2 & 1 \end{pmatrix}$$

for all $\lambda$.

If the characteristic of the sfield is not equal to 2, then the subgroup contains all the elements of the form

$$\begin{pmatrix} 1 & 0 \\ (\lambda + 1)^2 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ \lambda^2 & 1 \end{pmatrix}^{-1}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ 2\lambda & 1 \end{pmatrix}.$$

Therefore we have the theorem.

If the characteristic of the sfield equals 2, since

$$\begin{pmatrix} 1 & (\lambda + 1)^{-2} \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ \lambda^2 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ (1 + \lambda^{-1})^{-2} & 1 \end{pmatrix} = \begin{pmatrix} (\lambda + 1)^{-2} & 0 \\ 0 & (\lambda + 1)^2 \end{pmatrix},$$

the subgroup contains elements of the form

$$\begin{pmatrix} a^2 & 0 \\ 0 & a^{-2} \end{pmatrix}.$$

for all $a$. Let $a$ be any element of the sfield. From

$$\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a^2 & 0 \\ 0 & a^{-2} \end{pmatrix}\begin{pmatrix} 1 & \lambda^2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}^{-1}\begin{pmatrix} a^{-2} & 0 \\ 0 & a^2 \end{pmatrix} = \begin{pmatrix} 1 & a^2\lambda^2 a^2 + a^5 + a \\ 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} a^2 & 0 \\ 0 & a^{-2} \end{pmatrix}\begin{pmatrix} 1 & \lambda^2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a^2 & 0 \\ 0 & a^{-2} \end{pmatrix}^{-1}\begin{pmatrix} (a + 1)^2 & 0 \\ 0 & (a + 1)^{-2} \end{pmatrix} \times \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}\begin{pmatrix} (a + 1)^2 & 0 \\ 0 & (a + 1)^{-2} \end{pmatrix}^{-1},$$

we deduce that $\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$ belongs to the subgroup. Since $\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$ and its conjugates generate $SL_2(K)$ as $a$ runs over all elements of $K$, the theorem follows.

The previous method can be used to establish the easier

THEOREM 12. *Every normal subgroup of the general linear group $GL_n(K)$, which is not contained in the center, contains $SL_n(K)$, except the cases when $n = 2$ and $K$ has 2 or 3 elements.*

---

[1] Cartan, H., *Ann. école normale superieure*, **64**, 59–77 (1947), Theorem 4.

[2] Dieudonnè, J., *Bull. Soc. math. France*, **71**, 27–45 (1943).

[3] For the definition and properties of $PSL_n(K)$, see Dieudonnè.[6]

[4] Dickson, L. E., *Linear Groups*, Leipzig, 1901.

[5] Van der Waerden, B. L., *Gruppen von linearen Transformationen*, Berlin, 1935, p. 7.

[6] Iwasawa, M., *Proc. Imp. Acad. Japan*, **17**, 57 (1941).

[7] The author had some difficulty in understanding Dieudonné's proof. In fact, all the parabolic elements of $PSL_2(K)$ do not form a single conjugate set in $PSL_2(K)$.

---

# IMPRIMITIVITY FOR REPRESENTATIONS OF LOCALLY COMPACT GROUPS I

BY GEORGE W. MACKEY*

HARVARD UNIVERSITY AND UNIVERSITY OF CHICAGO

Communicated by Marshall Stone, July 21, 1949

*1. Introduction.*—In the classical theory of representations of finite groups by linear transformations a representation $s \to U_s$ of a finite group is said to be imprimitive if the vector space $H$ in which the $U_s$ act is a direct sum of independent subspaces $M_1, M_2, \ldots, M_n$ in such a manner that each $U_s$ transforms each $M_i$ into some $M_j$. In the present note we shall discuss a generalization of this notion which is more suitable for use in connection with infinite dimensional representation because it allows the direct sum decomposition to be continuous as well as discrete. Our principal theorem (well known for finite groups) deals with weakly (and hence strongly) continuous unitary representations of separable locally compact groups. It asserts that the pair consisting of such a representation and a "transitive system of imprimitivity" for it defines an essentially unique subgroup $G_0$ and an essentially unique representation $L$ of $G_0$ from which the original pair may be reconstructed in a quite explicit manner. This result has a number of applications. A recent theorem[1] of the author which implies the Stone-von Neumann theorem on the uniqueness of operators satisfying the Heisenberg commutation relations is included as a special case. In addition it may be used to give a complete determination of the irreducible unitary representations of the members of a class of locally compact groups which are neither compact nor Abelian.

2. *Definition of Imprimitivity.*—Let $s \rightarrow U_s$; $M_1, M_2, \ldots, M_n$ be an imprimitive representation in the classical sense. Suppose that the $U_s$ are unitary and that the $M_i$ are mutually orthogonal. Let $M$ denote the set of integers $1, 2, \ldots, n$. For each $s$ in the group $G$ and each $j \in M$ let $(j)s$ be the index of the subspace into which $U_{s^{-1}}$ carries $M_j$. Let $P_j$ denote the projection of $H$ on $M_j$. Then it is easy to see that $U_sP_jU_s^{-1} = P_{(j)s^{-1}}$. More generally if $P_E$ is defined by the equation $P_E = \sum_{(j \in E)} P_j$ for each $E \subseteq M$ then $U_sP_EU_s^{-1} = P_{(E)s^{-1}}$. The motivation for the following definition should now be clear. Let $M$ be a separable locally compact space and let $G$ be a separable locally compact group. Let $x, s \rightarrow (x)s$ denote a mapping of $M \times G$ onto $M$ which is continuous and is such that ($a$) for fixed $s$, $x \rightarrow (x)s$ is a homeomorphism and ($b$) the resulting map of $G$ into the group of homeomorphisms of $M$ is a homomorphism. Let $P(E \rightarrow P_E)$ be a $\sigma$ homomorphism of the $\sigma$ Boolean algebra of all Borel subsets of $M$ into a $\sigma$ Boolean algebra of projections in a separable Hilbert space $H$ such that $P_M$ is the identity $I$. Let $U(s \rightarrow U_s)$ be a representation of $G$ in $H$; that is a weakly (and hence strongly) continuous homomorphism of $G$ into the group of unitary operators in $H$. If $U_sP_EU_s^{-1} = P_{(E)s^{-1}}$ for all $E$ and $s$ and if $P_E$ takes on values other than $O$ and $I$ we shall say that $U$ is imprimitive and that $P$ is a system of imprimitivity for $U$. We shall call $M$ the base of $P$. It is to be observed that $P$ defines in $M$ a family of null sets and that there exists in $M$ a family of mutually equivalent measures whose sets of measure zero are precisely these null sets. The null sets are those sets $E$ for which $P_E = 0$ and the measures are those of the form $\mu(E) = (P_Ef, f)$ where $f$ is an element[2] in $H$ such that $P_Ef = 0$ implies $P_E = 0$.

3. *Ergodicity and Transitivity.*—When for each $x$ and $y$ in $M$ there exists $s$ in $G$ for which $(x)s = y$ it is natural to say that $P$ is a transitive system of imprimitivity for $U$. When $M$ is finite every system of imprimitivity decomposes in a natural manner into transitive ones corresponding to the orbits of $M$ under $G$. In general, however, the decomposition of $M$ into orbits is not reflected in a corresponding decomposition of $H$. It is rather the decomposition of $M$ into ergodic or metrically transitive parts which is relevant. Accordingly we define a system of imprimitivity $P$ to be ergodic if $G$ acts ergodically on the base $M$ of $P$; that is, whenever $(E)s$ differs from $E$ by a null set for all $s \in G$ then $E$ is itself a null set or the complement of one. In view of the current literature on the decomposition of measures the study of general systems of imprimitivity may be expected to be reducible to the study of ergodic systems.

Ergodic systems which are not also transitive are rather difficult to handle and such results as we have at present are far from definitive. This note will deal exclusively with transitive systems. Fortunately in some applications it can be shown that only transitive systems can arise. Specifically let us say that the orbits of $M$ under $G$ are regular if there exists a countable family $E_1, E_2, \ldots$ of Borel subsets of $M$, each a union of orbits such that each orbit of $M$ is the intersection of the members of a subfamily $E_{n_1}, E_{n_2}, \ldots$. Then the following theorem is easily proved.

THEOREM 1. *If the orbits of $M$ under $G$ are regular then for each ergodic system of imprimitivity based on $M$ there is an orbit $C$ such that $P_{M-C} = 0$.*

4. *Formulation of the Principal Theorem.*—Let $P$ be a transitive system of imprimitivity for the representation $U$ of the separable locally compact group $G$. Let $x_0$ be a point of the base $M$ of $P$. Let $G_0$ be the set of all $s \in G$ for which $(x_0)s = x_0$. Then $G_0$ is a closed subgroup of $G$ and the mapping $s \rightarrow (x_0)s$ of $G$ on $M$ defines a one-to-one Borel set preserving map of the homogeneous space $G/G_0$ of right $G_0$ cosets onto $M$. Thus $P$ is equivalent in an obvious sense to another system of imprimitivity for $U$ whose base is the homogeneous space $G/G_0$. In general we shall define a pair to be a unitary representation for the group $G$ together with a particular system of imprimitivity for this representation. If $U, P$ and $U', P'$ are two pairs with the same base $M$ we shall say that they are unitary equivalent if there exists a unitary transformation $V$ from the space of $U$ and $P$ to the space of $U'$ and $P'$ such that $V^{-1}U_s'V = U_s$ and $V^{-1}P_E'V = P_E$ for all $s$ and $E$. It follows from the above remarks that the problem of determining to within unitary equivalence all pairs based on a given $M$ may always be reduced to the corresponding problem in which $M$ is a homogeneous space. We shall accordingly confine ourselves to this case. The arbitrariness in the choice of $x_0$ has the effect only of providing us with several essentially equivalent complete systems of invariants for the pairs based on a given $M$.

Preparatory to stating our theorem we describe a method (which will prove to be general) of constructing pairs based on a given $G/G_0$. Let $\mu$ be a finite Borel measure on $G/G_0$ which is "quasi invariant" in the sense that the action of $G$ on $G/G_0$ preserves null sets.[3] Let $L(\xi \rightarrow L_\xi)$ be a representation of $G_0$ by unitary operators in a Hilbert space $H_0$. Then let $H_L$ be the set of all functions $f$ from $G$ to $H_0$ such that: ($a$) $f$ is a Borel function in the sense that $(f(s), v)$ is a Borel function of $s$ for all $v \in H_0$; ($b$) for all $s \in G$ and all $\epsilon \leftarrow G_0$, $f(\xi s) = L_\xi f(s)$; and ($c$) $(f(s), f(s))$ (which by ($b$) is constant on the right $G_0$ cosets) defines a summable function on $G/G_0$. By a more or less obvious adaptation of the proof of the Riesz Fischer theorem[4] it may be shown that $H_L$ is a Hilbert space with respect to the inner product $(f, g)_L = \int_{G/G_0}(f(s), g(s))d\mu$ and the obvious linear operations. Naturally functions which are equal almost everywhere are to be identified. Now let $\rho$ be the function on $G \times G/G_0$ which for each fixed $s$ is the Radon Nikodym derivative of the translate of $\mu$ by $s$ with respect to $\mu$ itself. Then regarding $\rho$, as we may, as a function on $G \times G$ let $U_s f$ for all $s \in G$ and $f \in H_L$ be defined by the equation $(U_s f)(t) = f(ts)/\sqrt{\rho(s^-, ts)}$. It is readily verified that $U_s$ is a unitary transformation of $H_L$ onto itself and that the mapping $s \rightarrow U_s$ is a representation of $G$. For each Borel subset $E$ of $G/G_0$ let $\phi$ be its characteristic function regarded as a function on $G$. For $f \in H_L$ let $(P_E f)(t) = \phi(t)f(t)$. It is easy to see that the mapping $f \rightarrow P_E f$ is a projection and that $U$ and $P$ together constitute a pair in the sense of the above definition. We shall call it the pair generated by $L$ and $\mu$. We can now formulate our main theorem.

THEOREM 2. *Let $G$ be a separable locally compact group and let $G_0$ be a closed subgroup of $G$. Let $U'$, $P'$ be any pair based on $G/G_0$. Let $\mu$ be any quasi invariant measure in $G/G_0$. Then there exists a representation $L$ of $G_0$ such that $U'$, $P'$ is unitary equivalent to the pair generated by $L$ and $\mu$. If $L$ and $L'$ are representations of $G_0$ and $\mu$ and $\mu'$ are quasi invariant measures in $G/G_0$ then the pair generated by $L'$ and $\mu'$ is unitary equivalent to the pair generated by $L$ and $\mu$ if and only if $L$ and $L'$ are unitary equivalent representations of $G_0$.*

5. *Outline of Proof.*—We shall give the proof in outline only leaving relatively routine details to the reader. Moreover we shall assume familiarity on the part of the reader with the paper cited in reference 1 and will omit arguments similar to those given there. We shall refer to this paper as SVN. The proof falls naturally into two parts. First we show that every pair defines a representation of $G_0$ unique to within unitary equivalence and that two pairs defining equivalent representations of $G_0$ are unitary equivalent. Then we complete the proof by showing that the representation of $G_0$ defined by the pair generated by an arbitrary $L$ and $\mu$ is unitary equivalent to $L$ itself.

Given a pair $U',P'$ based on $G/G_0$ we note first that the set of all $P_E'$ is a uniformly $n$ dimensional Boolean algebra of projections ($n = 1, 2, \ldots, \infty$) in the sense of Nakano (see SVN 5). This follows from the fairly easily proved fact that $G$ acting on $G/G_0$ is ergodic. Let $N$ denote an $n$ dimensional identity representation of $G_0$, let $\mu$ be a quasi invariant measure in $G/G_0$ and let $W,P$ be the pair generated by $N$ and $\mu$. Just as in No. 6 of SVN it is possible to show that the pair $U',P'$ is unitary equivalent to the pair $U,P$ where $P$ comes from the pair $W,P$ above and $U$ is a suitable representation of $G$. We define $Q_s$ as $U_s W_s^{-1}$ and observe that $Q_s P_E = P_E Q_s$ for all $E$ and $s$. It follows as in SVN that there exists a weakly Borel function $Q^\sim$ from $G \times G$ to the group of unitary operators in the space $H_1$ in which the $N_s$ act such that for each $s$ in $G$ we have $(Q_s f)(t) = Q^\sim(s, t)f(t)$. The identity $Q^\sim(s_1 s_2, t) = Q^\sim(s_1, t)Q^\sim(s_2, ts_1)$ holding for almost all triples is established as in SVN and from it the existence of a weakly Borel function $B$ such that $Q^\sim(s, t) = B^{-1}(t)B(ts)$ almost everywhere. The fact that the functions in $H_N$ are constant on the right $G_0$ cosets implies that $Q^{\sim}(s, \xi t) = Q^{\sim}(s, t)$ for all $\xi \epsilon G_0$ almost everywhere in $s$ and $t$. This implies in turn that $B^{-1}(\xi t)B(\xi ts) = B^{-1}(t)B(ts)$ in the same sense or equivalently that $B(\xi ts)B^{-1}(ts) = B(\xi t)B^{-1}(t)$. In short for each $\xi \epsilon G_0$, $B(\xi t)B^{-1}(t)$ is almost everywhere equal to a certain constant operator $L_\xi$. A simple argument shows that $(L_\xi v_0, v_1)$ is of the form $\int_G \psi(t)(B(\xi t)v_2, v_1)$ for a dense set of $v_0$'s. Here $v_0$, $v_1$ and $v_2$ are elements in $H_1$ and $\psi$ is a continuous complex valued function vanishing outside of a compact subset of $G$. It follows readily that $(L_\xi v_0, v_1)$ is continuous in $\xi$ and, since $L_{\xi_1\xi_2} = L_{\xi_1}L_{\xi_2}$, that $L(\xi \rightarrow L_\xi)$ is a representation of $G_0$. Of course $L$ may depend upon the choice of $\mu$, the choice of the unitary map of the given Hilbert space on $H_N$ and the choice of $B$. However, the fact that any two $\mu$'s have the same null sets guarantees the lack of dependence of $L$ on $\mu$. As to the other possible dependencies note that a unitary map $X$ of $H_N$ on itself which commutes with all $P_E$ is defined by an equation of the form $Xf(t) = X(t)f(t)$ where $X(t)$ is a unitary operator on $H_1$ for each $t$ and $X(t)$ is a weakly Borel function of $t$. Moreover $X(\xi t) = X(t)$ for $\xi \epsilon G_0$. It is readily calculated that the effect on $Q$ of a transformation by $X$ is to replace it by $R$ where $R(s, t) = X^{-1}(t)Q(s, t)X(ts)$. Now if $C^{-1}(t)C(ts) = X^{-1}(t)B^{-1}(t)B(ts)X(ts)$ it follows that $B(t)X(t)C^{-1}(t)$ is (modulo null sets) independent of $t$. Thus for some constant operator $K$ we have $C(t) = KB(t)X(t)$ so that $C(\xi t)C^{-1}(t) = KB(\xi t)X(\xi t)X^{-1}(t)\cdot B^{-1}(t)K^{-1} = KL_\xi K^{-1}$. In short our original pair and in fact the unitary equivalence class to which it belongs determine $L$ to within unitary equivalence. Conversely a simple reversal of the argument shows that pairs leading to unitary equivalent $L$'s must be themselves unitary equivalent.

Now let $L'$ be an arbitrary representation of $G_0$ and let $U',P'$ be the pair generated by $L'$ and a quasi invariant measure $\mu$ in $G/G_0$. By the argument of the preceding paragraph there is a unitary map $V^{-1}$ of $H_{L'}$ on some $H_N$ such that $V^{-1}P_E'V = P_E$ where $W,P$ is as before the pair generated by $N$ and $\mu$ and $N$ is an identity representation of $G_0$ on a Hilbert space $H_1$. It is not difficult to show that there exists a weakly Borel function $V^{\sim}$ defined on $G$ whose values are operators from $H_1$ to the space $H_2$ in which $L'$ operates such that $(Vf)(t) = V^{\sim}(t)f(t)$. It follows from the fact that $V$ is unitary that $V^{\sim}(t)$ is unitary from $H_1$ *into* $H_2$ for almost all $t$ and it follows from the fact that $Vf \epsilon H_{L'}$ that for each $\xi \epsilon G_0$, $V^{\sim}(\xi t) = L_\xi'V^{\sim}(t)$ for almost all $t$. Now the $Q_s$ of the preceding paragraph here take the form $V^{-1}U_s'VW_s^{-1}$ so that $U_s'VW_s^{-1} = VQ_s$. Hence $V^{\sim}(ts) = V^{\sim}(t)Q^{\sim}(s, t)$ or $V^{\sim}(ts) = V^{\sim}(t)B^{-1}(t)B(ts)$ or $V^{\sim}(ts)B^{-1}(ts) = V^{\sim}(t)\cdot B^{-1}(t)$. Thus there exists a norm preserving operator $K$ independent of $t$ such that $V^{\sim}(t) = KB(t)$ for almost all $t$. If $K$ were known to map $H_1$ onto the whole of $H_2$ we could write $B(t) = K^{-1}V^{\sim}(t)$ and conclude at once that $B(\xi t)B^{-1}(t) = K^{-1}L_\xi'V^{\sim}(t)V^{\sim -1}(t)K = K^{-1}L_\xi K$ and hence that the $L$ for $U',P'$ is unitary equivalent to $L'$. In order to show that $K$ is indeed an onto mapping we must make use of certain facts about the space $H_{L'}$ which so far as we know at this point might be zero dimensional. For each continuous function $w$ from $G$ to $H_2$ which vanishes outside of a compact subset of $G$ let $\tilde{w}$ be defined by the equation $(\tilde{w}(t), v) = \int_{G_0} (L_\xi'^{-1}w(\xi t), v)d\xi$ for all $v$ in $H_1$ and all $t$ in $G$. This function may be shown to be a continuous member of $H_{L'}$ which vanishes outside of a set whose image in $G/G_0$ is compact. Arguments of a fairly routine nature show that for each $t \epsilon G$ the vectors $\tilde{w}(t)$ span $H_2$. Now suppose that $K$ does not map $H_1$ onto $H_2$. Choose $v_0$ orthogonal to the range of $K$. Consider an arbitrary member of $H_{L'}$ of the form $\tilde{w}$. We have $\tilde{w}(t) = V^\sim(t)f(t)$ for some $f$ and almost all $t$. But $V^\sim(t)f(t)$ is in the range of $K$ for almost all $t$. Thus, since $\tilde{w}$ is continuous we can conclude that $(\tilde{w}(t), v_0) = 0$ for all $t$. Hence for all $t$, $(\tilde{w}(t), v_0) = 0$ for all $\tilde{w}$ and this contradicts the fact that the $\tilde{w}(t)$ span for each $t$.

6. *Reducibility.*—The natural question concerning the connection between the reducibility of a pair $U,P$ and the reducibility of the defining representation of $G_0$ is easily answered. If $T$ commutes with all $L_\xi$ then a transformation $T^\sim$ taking $H_N$ into $H_N$ is defined by the equation $(T^\sim f)(t) = B^{-1}(t)TB(t)f(t)$ where $B$ is the function used in defining $L$. Then, as is easily seen $T \rightarrow T^\sim$ is a $_*$-isomorphism of the ring of all bounded linear operators which commute with all the $L_\xi$ onto the ring of all bounded linear operators which commute with all the $U_s$ and all of the $P_E$. In particular the $U_s$ and the $P_E$ are simultaneously reducible if and only if $L$ is a reducible representation of $G_0$.

7. *Application to the Determination of Group Representations.*—Let $G$ be a separable locally compact group and let $G_1$ be a closed normal Abelian subgroup of $G$. Let $\hat{G}_1$ denote the character group of $G_1$. Every member $s$ of $G$ defines an automorphism $x \rightarrow sxs^{-1}$ of $G_1$ and this in turn induces an automorphism $y \rightarrow (y)s$ of $\hat{G}_1$. Now let $U$ be any irreducible representation of $G$. Restricted to $G_1$ it admits a spectral resolution defined by a $\sigma$ homomorphism $P$ of the Borel subsets of $\hat{G}_1$ into a Boolean algebra of projections in the Hilbert space $H$ in which $U$ acts. An obvious calculation shows that $U_sP_EU_s^{-1} = P_{(E)s^{-1}}$. Thus $P$ is a system of imprimitivity for $U$. Since $U$ is irreducible $P$ must be ergodic. If we assume that $G_1$ is "regularly imbedded" in $G$ in the sense that the orbits in $G_1$ under $G$ are regular then Theorem 1 tells us that $G_1$ may be replaced by a single orbit. Let $y$ be a point in this orbit and let $G_y$ be the closed subgroup of all $s$ for which $(y)s = y$. Theorem 2 tells us that $U$ is unitary equivalent to the first member of the pair generated by an irreducible representation of $G_y$.

If $G$ is a "semi-direct product" of $G_1$ and $G/G_1$; that is, if there exists a closed subgroup $G_2$ such that $G_1 n G_2 = e$ and $G_1G_2 = G$ much more precise information is available. The reader will have no trouble in verifying the truth of

THEOREM 3. *Let $G_1$ be imbedded regularly in $G$ and let $G$ be a semidirect product of $G_1$ and $G_2$. From each orbit $C$ of $G_1$ under $G_2$ choose a member $y_C$. Let $G_c$ denote the set of all $s \in G_2$ with $(y_C)s = y_C$. Then the general irreducible representation of $G$ may be obtained as follows. Select an orbit $C$ and an irreducible representation $L$ of $G_C$. Let $M$ be the irreducible representation of $G_1 \cdot G_C$ which coincides with $L$ on $G_C$ and is $y_C$ times the identity on $G_1$. Then the first member of the pair generated by $M$ and a quasi invariant measure in $G/(G_1 \cdot G_C)$ is the required irreducible representation of $G$. Every irreducible representation of $G$ may be so obtained and two such are unitary equivalent if and only if they come from the same orbit and unitary equivalent $L$'s.*

When the irreducible representations of $G_2$ and its subgroups are known Theorem 3 furnishes a complete description of the irreducible representations of $G$. This is so, in particular, when $G_2$ is Abelian. Moreover when $G_2$ is Abelian (and $G_1$ is imbedded regularly in $G$) it tells us that every irreducible representation of $G$ is of "multiplier" form. More generally any imprimitive representation of $G$ generated by a one-dimensional representation of a subgroup is unitary equivalent to a representation in which the underlying Hilbert space is the space of square summable functions on a homogeneous space and the action of the operator associated with $s$ is to translate by $s$ and multiply by a certain function (the multiplier) of $s$ and a variable point in the homogeneous space.

When $G_1$ is not imbedded regularly in $G$ Theorem 3 fails only in that it does not describe all of the irreducible representations. The ones that it does describe still exist and are irreducible. We have examples, however, showing that in general there are many others. Their existence leads to various kinds of pathological behavior which we expect to discuss at another time. Since these "extra" representations are all infinite dimensional, Theorem 3 provides an analysis of all finite dimensional representations for arbitrary semidirect products.

A number of well-known groups are regular semidirect products and Theorem 3 includes as special cases results in the literature analyzing their representations. Examples include[4] the unique non-commutative two-parameter Lie group[5] (a semidirect product of two lines) and the group of Euclidean motions in two space[6] (a semidirect product of the two-dimensional translation group and the circle group). Wigner's[6] reduction of the representation problem for the inhomogeneous Lorentz group to that for the homogeneous Lorentz group is also a consequence of Theorem 3 since the former group is a semidirect product of a translation group and the latter group.

8. *Concluding Remarks.*—In the mapping from a representation $L$ of $G_0$ to the pair it generates one can ignore $P$ and obtain a mapping from representations of $G_0$ to representations of $G$. It is not difficult to see that

this mapping carries the regular representation of $G_0$ into the regular representation of $G$. Thus (in view of No. 6) any analysis of the regular representation of $G_0$ as a direct sum or integral will define a corresponding analysis of the regular representation of $G$ although the "parts" will not necessarily be irreducible. This decomposition when $G_0$ is Abelian is the subject of a recent interesting note of Godement.[7] It was this note of Godement together with a discussion of such a space for compact groups given by A. Weil[8] that suggested our definition of the Hilbert space $H_L$.

There are a number of questions suggested by the considerations of this note which we expect to investigate and report on at a later date. We close by mentioning a few of these. (1) When is the representation $U$ of $G$ generated by an irreducible representation $L$ of $G_0$ itself irreducible? (2) When $G$ is finite, $L$ and $U$ are finite dimensional and $L$ is irreducible there is a classical theorem which says that the number of times that $U$ contains a given irreducible representation $V$ of $G$ is equal to the number of times that the restriction of $V$ to $G_0$ contains $L$. Weil[8] has recently extended this theorem to compact groups. One can ask whether (and in what sense) it continues to be true for general locally compact groups. (3) To what extent is it true that an arbitrary irreducible representation of $G$ is the imprimitive representation generated by a primitive representation $L$ of an appropriate $G_0$? How is the possible failure of this to hold generally connected with the "extra" representations of non-regular semidirect products? (4) Theorem 2 presumably can be used to prove other theorems like Theorem 3. What are some of these? One notes in particular that $G_1$ can probably be replaced by any group whose representations can be decomposed into irreducible parts in a suitably manageable manner.

---

[*] The author is a fellow of the John Simon Guggenheim Memorial Foundation on leave from Harvard University and in residence at the University of Chicago. A significant part of the work on this paper was done at each institution.

[1] Mackey, G. W., *Duke Math. J.*, **16**, 313–325 (1949).

[2] Cf. Nakano, H., *Ann. Math.*, **42**, 657–664 (1941), Hilfsatz 1.

[3] That such a measure always exists has been indicated by Dieudonné, *Ann. univ. Grenoble*, NS 23, 25–53 (1948) [p. 51]. Actually one can show, and this is important for our purposes, that any two quasi invariant measures have the same null sets. These null sets are precisely the sets whose inverse images in $G$ are null sets with respect to Haar measure.

[4] In carrying through the details of this and other arguments indicated in this note it seems to be necessary to know that every Borel set in $G$ which is a union of right $G_0$ cosets defines a Borel set in $G/G_0$. That this is so is an easy consequence of the following two theorems in the literature. A one-to-one continuous image of a Borel subset of a complete separable metric space is a Borel set. Kuratowski, C., *Topologie I*, Warsaw (1933), p. 251. If $g$ is a continuous function on a compact subset of a metric space then there exists a Borel subset on which $g$ is one-to-one and has the same range as before. Federer, H., and Morse, A. P., *Bull. Am. Math. Soc.*, **49**, 270–277 (1943) [Theorem 5.1].

[5] Gelfand, I., and Neumark, M., C. R. (Doklady), *Acad. Sci. Ukr. S. S. R.* (NS), **55**, 567–570 (1947).

[6] Wigner, E., *Ann. Math.*, **40**, 149–204 (1939).

[7] Godement, R., *C. R. Acad. Sci. Paris*, **228**, 627–628 (1949).

[8] Weil, A., *L'integration dans les groupes topologiques et ses applications*, Paris (1940), pp. 82, 83.

---

# *LINKAGE STUDIES OF THE RAT. X*

BY W. E. CASTLE AND HELEN DEAN KING

UNIVERSITY OF CALIFORNIA AND WISTAR INSTITUTE

Communicated August 1, 1949

1. *Fawn Linked with Agouti in Linkage Group V.*—In an earlier paper of this series, VIII, 1947,[1] a new color gene, fawn, was described and designated by the symbol, *f*. Preliminary linkage tests gave only negative results, but a cross between Agouti and fawn made by Dr. King indicated possible linkage between the two genes. Additional crosses of this nature made by Castle in Berkeley fully support that hypothesis.

A non-agouti black fawn stock, *aaff*, was crossed with a gray race, *AAFF*. The $F_1$ animals were gray of genotype $\frac{af}{AF}$. They were back-crossed to the black fawn stock, yielding four classes of young, expected to be equal numerically, unless linkage exists between *A* and *f*. The four genotypes expected and their observed frequencies were as follows:

| GRAY FAWN | BLACK FAWN | GRAY | BLACK |
|---|---|---|---|
| *Aaff* | *aaff* | *AaFf* | *aaFf* |
| 245 | 310 | 358 | 294 |

The two back-cross classes, gray fawn and black, would derive from crossover (recombination) gametes produced by the $F_1$ parent, whereas black fawn and gray would derive from non-crossover gametes. Clearly the non-crossovers exceed the crossovers numerically, indicating linkage.

The crossover percentage is $44.6 \pm 0.97$, a highly significant statistical result.

We conclude that genes *A* and *f* lie in the same fifth chromosome of the rat. The linkage is loose but unmistakable.

Linkage systems previously demonstrated, as now understood are as follows:

| | | | | | |
|---|---|---|---|---|---|
| I. | *p* | *r* | *c* | *l* | *w* |
| | 0 | 20.5 | 21 | 24.3 | 66.3 |
| II. | *Sh* | *Cu* | *an* | *in* | *b* |
| | 0 | 4 | 14.3 | 28 | 49 |
| III. | *hr* | | *wo* | | |
| | 0 | | 40.3 | | |
| IV. | *k* | | *st* | | |
| | 0 | | 34.1 | | |

2. *Further Evidence of Linkage Between Red-Eyed Yellow and Waltzing in the Rat.*—In earlier papers evidence has been presented that linkage exists in the rat between albinism, pink-eyed yellow and waltzing, all of which genes lie in the first chromosome of the rat[2]. Since red-eyed yellow is linked with pink-eyed yellow and with albinism, it follows that linkage should also exist between red-eyed yellow and waltzing. An opportunity to test this hypothesis by experiment recently presented itself.

A cross was made between a red-eyed hooded waltzer and a gray hooded non-waltzer, all the $F_1$ young being as expected, hooded and non-waltzers. A backcross to red-eyed waltzers produced as expected both red-eyed and dark-eyed waltzers but in very unequal proportions.

It is known that waltzing has very low penetrance, i.e., waltzers are regularly produced in smaller numbers than expected. In the backcross under discussion, if waltzing is a simple recessive, 50% of the young should be waltzers. In reality the percentage is much lower. In a total of 390 back-cross young, only 45 were recognizable as waltzers, this being 11.6% whereas 50% was expected.

Of the 45 waltzers, 30 were red-eyed and 15 dark-eyed, the non-waltzers numbering 180 red-eyed and 165 dark-eyed. In other words, the incidence of waltzing among the red-eyed was 1 in 7, but among the dark-eyed it was 1 in 12. This supports the idea that waltzing is linked with red-eye as well as with pink-eye and albinism, since otherwise red-eyed and dark-eyed waltzers should occur with equal frequency. In reality they deviate significantly from the equality expected, if linkage does not exist. Deviation, 7.5/P.E., 2.26 = 3.3.

*Acknowledgment.*—Grateful thanks are due to the Division of Genetics, College of Agriculture, University of California, and particularly to Dr. E. R. Dempster for facilities which made this investigation possible.

[1] Castle, W. E., and King, Helen Dean, "Linkage Studies of the Rat. VIII. Fawn, a New Color Dilution Gene," *J. Heredity*, **38**, 343 (1947).

[2] Castle, W. E., "Linkage in the Albino Chromosome of the Rat," PROC. NATL. ACAD. SCI., **32**, 33–36 (1946).

PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES
Volume 35 October 15, 1949 Number 10

## *NATIONAL ACADEMY OF SCIENCES*

# *THE CONFERENCE ON BIOELECTRIC POTENTIALS*

BY DUNCAN A. MACINNES

ROCKEFELLER INSTITUTE FOR MEDICAL RESEARCH

Communicated July 8, 1949

The first of a series of conferences on various rapidly developing subjects in science, under the Auspices of the National Academy of Sciences, was that on Bioelectric Potentials. It was held at the Rockefeller Institute for Medical Research in New York City, on December 14, 1946, in honor of Dr. W. J. V. Osterhout.

The discussion during the two-day session centered around four papers which were:

"Some Bioelectrical Problems," by W. J. V. Osterhout.
"The Source of Bioelectric Potentials in Large Plant Cells," by Lawrence Blinks.
"Physical Aspects of Bioelectric Phenomena," by Kenneth S. Cole.
"Nerve Action," by R. Lorente de Nó.

The first three of these papers, revised in the light of the discussions, follow in this number of the PROCEEDINGS.

The participants in the conference were: L. R. Blinks, Detlev W. Bronk, Kenneth S. Cole, Richard T. Cox, P. Debye, Joseph Erlanger, Teh-Pei Feng, W. O. Fenn, R. H. Fuoss, Herbert S. Gasser, H. S. Harned, R. Höber, J. G. Kirkwood, Irving Langmuir, David P. C. Lloyd, L. G. Longsworth, D. A. MacInnes, R. Lorente de Nó, L. Michaelis, John von Neumann, Lars Onsager, W. J. V. Osterhout, Alexandre Rothen, Francis O. Schmitt and Theodore Shedlovsky.

The other conferences that have been held so far are as follows:

| *Subject* | *Chairman* | *Date* |
|---|---|---|
| Foundations of Quantum Mechanics | K. K. Darrow | June 2–4, 1947 |
| Theoretical Physics | J. R. Oppenheimer | March 30 to April 2, 1948 |
| Low Temperature Physics | J. C. Slater | May 31, June 1–2, 1948 |
| The Gene | B. P. Kaufmann | May 31, June 1–2, 1949 |
| The Ultracentrifuge | D. A. MacInnes | June 12–14, 1949 |

These conferences were held at the Rams Head Inn, Shelter Island, L. I., N. Y., except the second which took place at Pocono Manor, Pa.

The conferences have been organized so as to bring together active workers in order to consider current research problems in a selected field of science. Formal presentation of papers has been avoided, the emphasis being on leisurely and informal discussion. The number of participants in each conference has been limited to 25. After the first conference, which was held in New York City, the remaining ones have been held at country inns.

It has been found that much of value may be accomplished in two or three days of close association by key workers, who are frequently widely scattered geographically. Concentration on a given topic, the absence of distracting interests and the opportunity of free intimate discussion which is made possible by a small group, are particularly helpful.

These conferences have been made possible by the advice and encouragement of Past President Jewett and President Richards, both of whom assigned Academy funds for the expenses involved.

---

# *SOME BIOELECTRICAL PROBLEMS*

By W. J. V. OSTERHOUT

THE LABORATORIES OF THE ROCKEFELLER INSTITUTE FOR MEDICAL RESEARCH

Communicated July 8, 1949

What would Galvani say if he could join us today? Perhaps the conversation would run as follows.

"In 1791 I published some observations on frogs' legs. I noticed that in contact with copper and iron they twitched vigorously. Can you tell me why this happens?"

"We fear that we cannot give you a satisfactory answer. In spite of much study the cause is not clear."

"It is a long time," he answers, "since the first observations were made."

"It is true, but we need more help from the physicists and physical chemists to solve this problem. That is the reason for this conference."

"Well," he says finally "how far have you got?"

A clear and simple answer is not easy to give. Much progress has been made but much remains obscure. For the purpose of our conference it may be best to consider some fundamental questions which can be so stated as to appeal to the physicist and to the physical chemist.

If I attempt a brief statement of this kind as requested by our chairman I must ask your indulgence.

The behavior which astonished Galvani is due to disturbances in the electrical potentials which exist at the surfaces of living cells. We can measure these potentials and the rate at which these disturbances travel along the cell and we can subject our observations to mathematical treatment. We have made some progress in analyzing the mechanism by which an electrical disturbance is produced and propagated. In some cases the potentials appear to be due chiefly to diffusion potentials set up by potassium compounds against the non-aqueous surface films of the protoplasm. In such cases we can measure the mobilities of the ions in the non-aqueous films and show that they account for the observed potentials. These mobilities are usually very different from those found in water.

If a spot with normal resting potential is connected through a galvanometer to another spot on the same cell where the resting potential has been reduced to zero (by applying KCl) we find in some cases that a current of several microamperes flows from the normal to the treated spot and this may persist for several hours. It would appear that phase boundary potentials as ordinarily understood could not account for such currents.

To illustrate some of the methods employed in such studies we may consider experiments with a plant cell, *Nitella*, which acts somewhat like muscle and nerve but is in some respects easier to study.

The protoplasm of the *Nitella* cell (Fig. 1) forms a thin layer (not over 15 microns thick) surrounding a large central vacuole (about 500 microns in diameter) filled with sap; outside the protoplasm is a cellulose wall which is so permeable that it does not affect the potentials (Fig. 1). We can obtain sap from the interior of such cells without external contamination and we can follow changes in its composition by expressing and analyzing it.

*Surface Films.*—The surfaces (both the outer surface and the inner surface surrounding the vacuole) of the protoplasm are completely covered with a film of oily or non-aqueous material too thin to be visible under the microscope and hence less than half a wave-length of visible light in thickness. We do not know whether these films are solid or liquid or an emulsion but they act like liquid in obeying the laws of surface tension.

Although protoplasm is largely composed of water it rounds up in contact with water like an oily liquid because of the non-aqueous film covering its surface. This is easily seen if the cellulose wall is pierced with a needle so that the protoplasm protrudes into water or if the protoplasm is withdrawn from the cellulose wall by a plasmolyzing solution which fills the space between the protoplasm and the cellulose wall. The inner protoplasmic surface acts in similar fashion as is evident when it is rapidly deformed by protoplasmic motion.

It is natural to assume that such films are formed by substances which lower the surface tension of the protoplasm and which therefore accumulate in the surface since, if they diffuse into it, energy is required to get them out. We might therefore expect the inner and outer protoplasmic surfaces to be similar. This is approximately true of *Nitella* for if we place sap outside the cell and form the chain

Sap | Protoplasm | Sap

we find an e. m. f. of only 16 mv.[1]

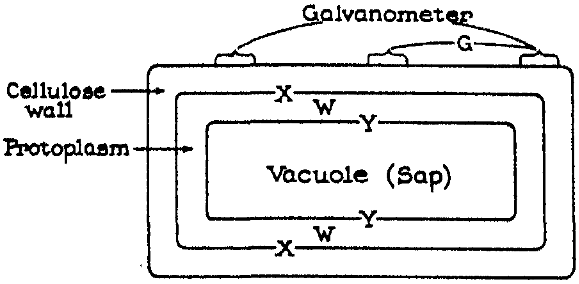


FIGURE 1

Diagram to represent the structure of a *Nitella* cell. The protoplasm consists of an aqueous layer, $W$, with a film of non-aqueous material, $X$, at the outer surface and a corresponding film, $Y$, at the inner surface. The vacuole is filled with sap containing about 0.05 $M$ KCl and 0.05 $M$ NaCl with a very small amount of other electrolytes and of organic matter. (The relative diameter of the vacuole is much greater than is here represented.)

These films determine what substances enter and leave the cell and thus control metabolism. They furnish the best test of life which we possess since if they are irreversibly destroyed life ceases at once. This is indicated by the entrance of such dyes as trypan blue or acid fuchsin which do not enter the cell as long as the surface films are intact and which thus provide a useful test of death. We may say that the cell is dead when the protoplasm becomes completely and irreversibly permeable as happens when the surface film is destroyed.

The destruction of the surface film may be brought about by agents which act on the film or on the substrate of protoplasm on which the film rests.

The bioelectrical potentials are set up at these films. Since the study of such films involves surface chemistry and physics which are little understood it is highly desirable to find suitable models for study.

Attempts have been made to produce such films between two aqueous solutions to imitate the conditions in the living cell but so far the results have not been satisfactory. One method is to fill a tube with aqueous solution and place it upright in a non-aqueous liquid and then allow a drop of aqueous solution hanging from another tube to descend slowly until it compresses the non-aqueous liquid between the two aqueous solutions. In this way very thin non-aqueous films can be obtained and electrical measurements across them can be carried out, but the films soon disrupt. Films have also been made by blowing bubbles with non-aqueous materials but they do not last long.

*Resting Potentials.*—By means of its surface films the cell can act very much like a potassium electrode. To see this we need only connect a galvanometer between a spot on the cell where the potential is kept constant to another spot which is in contact with an aqueous solution of KCl. As the concentration of KCl at the latter spot changes while that at the other spot remains constant the P.D. changes in such fashion as to obey the equation for diffusion potential which for 25°C. may be written

$$\text{P.D.} = 59 \frac{u_{\mathrm{K}} - v_{\mathrm{Cl}}}{u_{\mathrm{K}} + v_{\mathrm{Cl}}} \log \frac{a_1}{a_2}$$

Here $u_{\mathrm{K}}$ and $v_{\mathrm{Cl}}$ are the mobilities of $\mathrm{K}^+$ and of $\mathrm{Cl}^-$ in the outer non-aqueous protoplasmic surface film, $X$; $a_1$ and $a_2$ are the activities[2] in $X$ of $\mathrm{K}^+$ (taken as equal to those of $\mathrm{Cl}^-$) when two solutions are applied to $X$ in succession so as to set up a diffusion potential in $X$. We find that $u_{\mathrm{K}}$ may be 73 times $v_{\mathrm{Cl}}$ which shows how much the surface of the protoplasm differs from water in which $u_{\mathrm{K}} \div v_{\mathrm{Cl}} = 0.96$.

To carry out such measurements we allow the cell to remain for an hour or more in contact with 0.001 *M* KCl. We assume that the outermost part of $X$ has reached approximate equilibrium with the external solution so that $a_2 = 0.001\ (S)$ where $S$ is the activity partition coefficient of $\mathrm{K}^+$ (taken as equal to that of $\mathrm{Cl}^-$).[3] We then replace 0.001 *M* KCl by 0.01 *M* KCl and assume that a diffusion gradient between the two activities of $\mathrm{K}^+$ is set up in $X$ and that $a_1 = 0.01\ (S)$. Since the value of $S$ remains constant[4] as long as the properties of the two phases ($X$ and the external solution) remain constant a tenfold change in the activity of $\mathrm{K}^+$ in the external solution means a tenfold change in the activity of $\mathrm{K}^+$ in $X$.

When we set up models and use various non-aqueous substances in place of $X$ we find that a tenfold change of activity of KCl in the aqueous solution gives different values for the P.D. depending on the nature of the non-aqueous phase. This must be due to differences in $u_{\mathrm{K}}$ and $v_{\mathrm{Cl}}$ if the activity partition coefficient $S$ is constant.

The situation may be more complicated if other ions in the surface play an important part, but for our present purpose they may be left out of consideration since as a rule the effect in *Nitella* is chiefly due to KCl.

It may be asked whether the surface behaves like a system of charged pores such as exist in a collodion membrane. This is evidently not the case for in *Nitella* we find that under some conditions the order of mobilities in $X$ is $u_K > v_{OH} > u_{Na}$ and in some other plants it is $u_K > v_{Cl} > u_{Na}$. In a system of charged pores all cations are faster than all anions or vice versa.

By means of its surface films the protoplasm is able to distinguish clearly between NaCl and KCl. When 0.01 $M$ KCl in contact with *Nitella* is replaced by 0.01 $M$ NaCl the outwardly directed potential is increased by about 80 mv. This has been called the potassium effect.

Some substance (or group of substances) in the protoplasmic surface must be responsible for this and a number of materials have been examined from this point of view. The most interesting so far tested is nitrobenzene. When two tubes filled with aqueous solutions dip into nitrobenzene, one containing 0.1 $M$ KCl and the other 0.1 $M$ NaCl, we find on connecting them through a galvanometer that 0.1 $M$ NaCl is positive to 0.1 $M$ KCl just as in *Nitella*. In the living cell the value is about 80 mv. and in nitrobenzene about 67 mv.

It would be very interesting to know how nitrobenzene produces this result. We hope that some physical chemist will look into this.

The non-aqueous films at the surfaces of the protoplasm are easily altered. Leaching *Nitella* with distilled water reduces the relative mobility of $K^+$ in the outer protoplasmic surface so that $u_K \div u_{Na}$ falls from 33 to 1 and $u_K \div v_{Cl}$ falls from 73 to 2.2. Hence the cell is no longer able to act like a potassium electrode or to distinguish electrically between $K^+$ and $Na^+$. At the same time it loses its power to respond to an electrical stimulus. Evidently a substance or group of substances is removed from the surface films; it seems possible that such substances are found in blood for if we apply dilute blood to the leached cell for a few seconds we restore its ability to distinguish between $K^+$ and $Na^+$ and to respond to electrical stimulation.

This important substance exists in very minute amounts and we need the help of the chemist to study it. A beginning has been made by taking water in which *Nitella* has been standing and shaking it with chloroform. A precipitate then separates out at the surfaces of the drops of chloroform. When this is redissolved in water and applied to the cells it restores them to their normal state. But the amount of substance thus obtained is very small and we need better methods.

Interesting alterations in the surface films are produced by guaiacol. In *Nitella* it alters mobilities and in *Valonia* it changes the order $u_K > v_{Cl} > u_{Na}$ to $u_{Na} > v_{Cl} > u_K$.

It is evident that changes in the surface films due to metabolism might alter mobilities. A variety of changes may occur under normal conditions in the surface films. This is not surprising for any substances produced in the cell which have the power to lower surface tension become trapped if they enter the surface since energy is required to get them out. Partition coefficients may also be altered so that without any change in the concentration of NaCl or of KCl in the external solution their relative proportions in the non-aqueous surface film may change and thus affect the potential.

Since the protoplasmic surface acts like guaiacol in some respects a study of partition coefficients in guaiacol has been made by Shedlovsky and Uhlig which is very helpful in dealing with the living cell.

The partition coefficients in *Nitella* may be calculated as follows. Putting for convenience $v_{Cl} = 1$ and ascertaining the relative values of $u_K$ and $u_{Na}$, as already described (p. 550) we can predict the value to be expected in a given mixture of NaCl and KCl. Any divergence from the predicted value is then attributed to differences in the partition coefficients and these are evaluated. They are then used to predict values expected in other mixtures and if these prove to be approximately correct we assume that we have ascertained the approximate values of the partition coefficients. When we apply this same method to guaiacol in a U-tube used in place of *Nitella*, we get approximately correct values.

What determines the magnitude of the potentials in *Nitella*? Since the changes in *Nitella* obey the equations of Nernst and of Henderson for diffusion potential and these equations enable us to predict the values found in various mixtures of electrolytes we may suppose that we are dealing with diffusion potentials. We can determine the relative values of $u_K$ and $v_{Cl}$ for the outer protoplasmic surface in *Nitella* (p. 550) and the P.D. across the protoplasm can be calculated from the activity gradient of KCl across the protoplasm (since the other electrolytes present have relatively little effect on the P.D.).

The sap in contact with the inner protoplasmic surface contains about 0.05 *M* KCl and if we put 0.001 *M* KCl outside the protoplasm we may write for 25°C., if $u_K = 73$ when $v_{Cl}$ is taken as unity,

$$\text{P.D.} = 59 \frac{73 - 1}{73 + 1} \log \frac{0.05}{0.001},$$

where 0.05 and 0.001 are the concentrations (here used in place of activities) of KCl in the sap and in the external solution, respectively. This gives 97 mv. The average observed value is usually about 100 mv.

It would therefore appear that the diffusion potential of KCl in the sap is sufficient to account for the observed potential. On this basis we may expect to change the potential by changing the concentration of KCl in the sap. This can readily be done. If we place 0.24 *M* sucrose + 0.001 *M*

KCl at one end of the cell, called *A*, and 0.001 *M* KCl at another spot, designated as *B*, we see that water enters at *B*, moves through the cell and comes out at *A*. This produces a visible rush of particles in the vacuole from *B* to *A*. This movement of water carries KCl from *B* to *A* but KCl cannot pass through the protoplasm at *A* (except very slowly) and consequently it becomes more concentrated at *A* as water escapes into the external solution of sucrose. Hence the concentration of KCl increases at *A* and falls off at *B*. At the same time the potential increases at *A* and falls off at *B*. This can be shown by placing electrodes at *A* and *B* and connecting them to a third electrode, *C*, at another spot where the potential is kept constant and sap is prevented from moving.

When we replace the sucrose solution at *A* by 0.001 *M* KCl we see that water enters at *A*, moves through the cell and escapes at *B*, thus reversing the flow of sap. The concentration of KCl falls off at *A* and increases at *B* and in consequence the potential falls off at *A* and increases at *B*.

These experiments illustrate some of the methods used to determine the nature of bioelectric potentials. Although they have been developed especially in the study of large plant cells they can be applied to a considerable extent to other cells.

This study of resting potentials leads us to ask how these potentials are altered when an electrical disturbance travels along the cell. Since the preceding account has dealt with potentials in *Nitella* the simplest procedure is to continue with an account of electrical disturbances in this cell.

*Stimulation.*—This subject can not be discussed in detail but some of its fundamental features may be briefly mentioned. It consists of an electrical disturbance which is produced at one point and which usually travels along the cell.

*Chemical Stimulation.*—When the external solution is 0.001 *M* NaCl there is an outwardly directed potential of about 100 mv. at all points (Fig. 2). This is due to the diffusion potential of KCl at the non-aqueous film, *Y*, forming the inner protoplasmic surface. If we replace 0.001 *M* NaCl by 0.01 *M* KCl at *A* we create an inwardly directed potential of 100 mv. at the outer surface which balances the outwardly directed potential of 100 mv. at the inner surface and thus reduces the net potential at *A* to zero. At the same time the electrical resistance at *A* falls to about 1% of the normal, since both the inner and the outer surfaces of the protoplasm are in contact with a relatively high concentration of $K^+$ (the sap contains about 0.05 *M* KCl). All of these facts are established by direct measurement.

In this way a difference of potential between *A* and *B* is set up but if the cell is in a favorable state this difference soon decreases because the potential at *B* falls off in such fashion that *B* may be said to discharge into *A*. Then *C* discharges into *B* and in this way the disturbance travels along the cell. During the discharge we may suppose that KCl moves out of the sap into the external solution, thereby lowering the outwardly directed activity gradient of $K^+$ on which the diffusion potential depends. There may be an actual rupture or a great increase in the permeability of the protoplasmic surface films at *B* when the electrical pressure reaches a certain point. If a film is 0.1 micron thick as may well be the case and the potential is 100 mv. the electrical gradient across the film just before the discharge is 10,000 volts per centimeter.

*Electrical Stimulation.*—It might be expected, in view of these facts, that when the external solution is 0.001 *M* NaCl and the potential 100 mv. at all points an e. m. f. applied by means of two electrodes in such fashion as to create a sufficient outgoing current at *B* and an ingoing current at *A* might cause $K^+$ to move outward at *B* and reduce the potential at *B* so that the next spot, *C*, could discharge into it. This appears to be the case but it is necessary to apply more than 100 mv. since the resistance at *A* is now higher than in the former case. In describing chemical stimulation it was stated that the current entered *A* where it was in contact with 0.01 *M* KCl and the resistance was only 1% of that in a region in contact with 0.001 *M* NaCl, as in the present case.

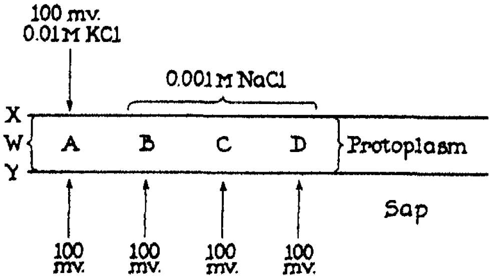


FIGURE 2

Diagram to illustrate the potentials assumed to be present in a *Nitella* cell. At the inner non-aqueous surface film, *Y*, there is a potential of 100 mv. at all points. At the outer film, *X*, no potential is assumed to exist when the external solution is 0.001 *M* NaCl but when this solution is replaced by 0.01 *M* KCl there is an inwardly directed potential of 100 mv. at *X* so that the net potential across the protoplasm is zero. No precise assumption is made regarding the concentration of KCl in the sap.

Stimulation occurs with an outgoing electrical current which carries $K^+$ outward, not with an ingoing current which carries $K^+$ inward.

When $K^+$ moves out of the sap this movement is opposed by a process which is continually operating to drive $K^+$ inward and to keep its concentration inside the cell much greater than outside. The following suggestion may be made in regard to the process.

The experiments indicate that ions at the outer surface of the non-aqueous layer (surrounding the protoplasm) may combine with organic substances formed only at this spot and may then migrate through this layer to the aqueous protoplasm. If the organic substances are then irreversibly decomposed so as to set the ions free it is evident that as free ions they cannot pass out, except very slowly, since they are only slightly soluble in the non-aqueous surface layer. Hence they will be trapped and accumulate inside. We may assume that the reaction which decomposes the organic substances yields energy.

*Recovery.*—When $K^+$ has moved outward at *B* and the potential at *B* has fallen $K^+$ is moved back into the vacuole by the accumulative process. This is called recovery.[5]

Recovery involves a return to the sap of $K^+$ and other electrolytes which have moved outward. It also requires restoration to normal of surface films which have suffered change.

If an impulse arrives before recovery is complete there may be a loss of whatever potential has been recovered at that time but the magnitude of the change falls short of the normal response. For a short time after stimulation no response is possible ("absolute refractory period") but as recovery proceeds the response increases with progressive restoration of the outwardly directed activity gradient and outwardly directed diffusion potential at *Y*.

*Propagation of Impulses.*—When the potential at *A* falls to a certain value so that the next spot, *B*, discharges into *A* we see that *B* becomes an electrical sink into which the next spot, *C*, discharges and in this way the disturbance travels along the cell. The rate of travel in *Nitella* at 25°C. is usually about 2 cm. per second but may be much faster. In any case it is very much slower than in muscle and nerve.

If the traveling impulse reaches a spot (known as a "block"), which is unable to respond by a discharge, the impulse can progress no further. If we have three adjoining spots, *A*, *B* and *C*, with *A* and *C* normal and *B* serving as a block, the impulse can travel past *B* only if we connect *A* and *C* by a moist thread through which *C* can discharge into *A*. The impulse can then continue to travel along the cell. A gap between two cells can be bridged over in the same way by means of two moist threads which complete the circuit so that the impulse can travel from one cell to the other.

A block may consist of an area of very low potential or it may be a region which has the normal resting potential but does not discharge. Normal cells of *Nitella* can be rendered inexcitable by anesthetics or by leaching in distilled water. The leaching apparently removes from the cell a substance (called *R* for convenience) which makes the cell excitable. It also ensures a high mobility for $K^+$ in the outer protoplasmic film and this produces the potassium effect. It can be extracted from the distilled water used for leaching and applied to the cell which then reacts normally. A substance which has the same restorative effect is found in blood and it seems possible that it plays an important part in maintaining the normal excitability of muscle and nerve.

*Mechanical Stimulation.*—Mechanical stimulation in *Nitella* presents a picture which differs from that hitherto described.

The most convenient way to produce mechanical stimulation is by pinching the cell. This sets up a compression wave which travels along the cell (presumably with the speed of sound in water) and causes a loss of potential at each spot it reaches. The effect may be reversible or irreversible depending on the violence of the disturbance.

If the cellulose wall is imbibed with 0.001 *M* NaCl there is an outwardly directed (positive) potential[6] everywhere and when the compression wave arrives at any point the loss of this positive potential causes a "negative variation." But if the cellulose wall is imbibed with 0.1 *M* KCl there is an inwardly directed (negative) potential at all points and when the compression wave arrives the loss of this negative potential causes a "positive variation."

It is evident that more investigation is needed at almost every point to clear up the mechanisms involved in stimulation. The energy relations are of interest since stimulation is accompanied by an increased output of $CO_2$ in nerve.

In conclusion it seems desirable to stress again the importance of the non-aqueous surface films. These structures, invisible under the highest magnification, play an all-important rôle. They are the seat of considerable electrical forces so that when a sufficient number of cells is in series an e. m. f. of 500 volts may be available, as in the electric eel. The high electrical capacity and resistivity of the cell is due to them. They regulate the intake and outgo of all substances and thus control metabolism. If they are destroyed death ensues at once. The behavior of such films involves the little understood laws of surface chemistry and physics and deserves intensive study.

The problems encountered in the study of *Nitella* are for the most part closely related to those we meet in experiments on muscle and nerve. Many other cells present different problems of equal interest and complexity. It is our hope that this conference may bring us more aid from physicists and physical chemists in dealing with these important questions.

---

[1] In some cases larger values are observed. In *Halicystis* we find 50 mv. The cause of this is not known.

[2] For convenience activities will be regarded as equal to concentrations in all cases in this paper.

[3] If the same standard state is assumed for the solution in both phases $S$ would be unity, but it is more convenient to assume different standard states so that as long as the properties of the two phases remain constant $S$ will be constant but as a rule its value will not be unity.

[4] The partition coefficient includes only the ions $K^+$ and $Cl^-$. It would not be constant if it included the undissociated KCl in $X$ (but this undissociated part would not affect the diffusion potential), cf. Osterhout, W. J. V., *J. Gen. Physiol.*, **26**, 293 (1942–1943).

[5] If sufficient $K^+$ is present in the external solution there is no recovery since the accumulative forces do not produce a sufficiently great outwardly directed activity gradient of $K^+$.

[6] I.e., cations tend to flow across the protoplasm from the sap to the external solution.

---

# *SOME PHYSICAL ASPECTS OF BIOELECTRIC PHENOMENA*

BY KENNETH S. COLE*

INSTITUTE OF RADIOBIOLOGY AND BIOPHYSICS, UNIVERSITY OF CHICAGO

Communicated by D. A. MacInnes, July 8, 1949

The structure and activity of living systems present many problems of physics, chemistry and mathematics and the electrical aspects of these problems have been investigated almost continuously for the past century and a half. As increasingly powerful concepts, techniques and instruments have been developed they have been applied to the study of these phenomena and, as a result, we now have available a vast store of information on the bioelectric phenomena of organisms, tissues and cells—both plant and animal. Much of this information has been correlated into the general principles of electrophysiology and, in some cases, in surprisingly successful quantitative form. But the further description and interpretation of the various bioelectric phenomena in terms of physical and chemical concepts has been, in general, less successful. Although it has been almost impossible to design and execute experiments on living materials which disclose their physical characteristics in simple and direct form, some of these characteristics have been measured and found to be strikingly similar to phenomena of non-living systems. This is highly gratifying to the physical scientist and very useful to the biologist except for the several important instances in which the behavior of these non-living systems has itself not been satisfactorily explained. Thus, as is found for other biological phenomena, the physical aspects of the electrical phenomena may present problems of experiment and theory which are not confined to the living systems alone but extend into the physics and chemistry of non-living liquid and solid materials.

This contribution to the Conference on Bioelectric Potentials presents some of the simpler electrical characteristics of living cell membranes which have come from the application of physical principles and techniques to the measurement and analysis of the electrical properties of cells and tissues. These membrane characteristics may serve the Conference both as background information and as examples of the many unsolved physical problems of biological systems.

Although the experimental methods used in these investigations of the electrical characteristics of the cell membranes cannot be discussed in detail they should not be ignored completely. The design of the experiments and the equipment and the analysis of the results have usually been rather straightforward applications of conventional techniques such as those of potential theory, circuit analysis, electrochemistry, and communication and electronic engineering. But as often as not the conditions imposed by living cells have required that the necessary techniques be pushed beyond those achieved in current practice and in some cases the equipment has had to reach to the natural limits of sensitivity and resolution.

*Membrane of the Cell.*—It is well established not only that the interior of the living cell is very different from the external inanimate environment in composition, structure and electric potential, but also that these differences are maintained by a barrier at the surface which is necessary for the life of the cell. Although this barrier may not be positively identified under the microscope—as for example a part of the plasma membrane—it is definitely recognized by other characteristics—such as a rather small permeability to water and to many solutes. The structure which constitutes the barrier to the free flows of ions in and out of the cell is most probably the seat of the immediate source of electrical energy and the origin of the principal bioelectric effects. This structure is our primary concern and for convenience we shall refer to it as the membrane of the cell.

*Capacity of the Membrane.*—The cell membrane was thought of electrically as a "leaky condenser" before 1900 and by 1925 the capacity of this condenser had been measured in the red blood cell[1] and found to be about one microfarad per square centimeter. Since that time the membrane capacity of an adequate number of cell types has been measured[2] with sufficient accuracy to justify the hypothesis that all living cells have a membrane capacity and that it is of the order of one microfarad per square centimeter. On the assumptions of a macroscopic behavior and a dielectric constant of three, the membrane has a thickness of 30 A. This thickness is clearly of molecular dimensions and it suggests the need for more detailed information and analysis on thin films. A further complication is that only a few of these membranes exhibit a pure static capacity, it being more usual to find a behavior which resembles the dielectric loss of engineering insulators.

The alternating-current characteristics of the membrane capacity are conveniently considered[3] in two parts—$C'$, the conservative or "real" component and $C''$ the dissipative or loss, component—which are combined as the complex capacity,

$$C^* = C' - iC'', \tag{1}$$

where $i = \sqrt{-1}$. The membrane capacities are adequately expressed empirically[4] by

$$C^* = \bar{C}(i\omega)^{-(1-\alpha)},$$

where $\bar{C}$ is constant, $\omega/2\pi$ is the frequency and $\alpha$ is a loss index, $0 \leq \alpha \leq 1$, which is related to the phase angle $\phi = \alpha\pi/2$. These results can be expressed in terms of a resistance and a capacity, both of which vary with frequency, or, as shown in figure 1, in terms of $C'$ and $C''$.

This behavior of cell membranes is similar to the polarization of electrode systems such as platinized platinum, silver–silver chloride and calomel, and this similarity has led to the attractive suggestion that this membrane capacity also is a phenomenon of ion transport. Although it may not be possible either to accept or reject the hypothesis without a better understanding of the electrode systems, several factors make it seem rather improbable:

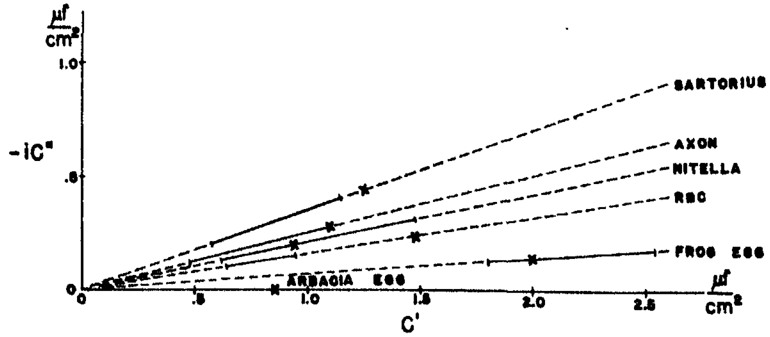


FIGURE 1

Complex capacity loci, C′ *vs.* C″, for various cell membranes; in each case the solid line represents the experimentally determined portion of the locus and X is the observed or calculated point for a frequency of 1 KC.

(*a*) This analytical description of the membrane and electrodes is not unique to them but can be applied with equal success to the electrical properties of dielectrics[5] and a barrier layer photovoltaic cell[6] and to the mechanical characteristics of some viscoelastic materials.[7]

(*b*) The capacities of the electrode systems are usually one or more orders of magnitude larger than those of the cell membranes.

(*c*) Although the electrode characteristics depend to a considerable degree upon the method of manufacture and the composition and concentration of the electrolyte, the membrane characteristics seem quite stable and remain relatively unchanged as the cell becomes active, is injured or is placed in a chemically modified medium.

(*d*) The anomalous reactance, to be mentioned later, has some characteristics which are more closely analogous to those found in ion transport processes than have these membrane capacities of one microfarad per square centimeter.

The membrane capacity may—as has been indicated—be dielectric in origin. For a dilute solution of dipoles the complex dielectric factor $\epsilon^*$ is given by

$$\epsilon^* = \epsilon - i\epsilon = \epsilon_\infty + (\epsilon_0 - \epsilon_\infty)/(1 + i\omega\tau), \tag{2}$$

where $\epsilon_0$ and $\epsilon_\infty$ are the dielectric constants at zero and infinite frequencies, respectively, and $\tau$ is the relaxation time.[8] An equivalent circuit is readily interpreted with one capacity for $\epsilon_\infty$, the deformation polarization, another capacity for $\epsilon_0 - \epsilon_\infty$, the rotation polarization, and a resistance for $\tau/(\epsilon_0 - \epsilon_\infty)$, the viscous dissipation in rotation. For concentrated solutions of dipoles and polar liquids and solids one often finds[5]

$$\epsilon^* = \epsilon_\infty + (\epsilon_0 - \epsilon_\infty)/[1 + (i\omega\tau)^{1-\alpha}]. \tag{3}$$

In the equivalent circuit the resistance is now replaced by an impedance $z^* = \bar{z}(i\omega)^{-\alpha}$. If now $\epsilon_\infty << \epsilon' << \epsilon_0$ and $\alpha$ is near unity the element $z^*$ is the only one of importance and equation (3) approaches the membrane characteristic of equation (1) and figure 1. Then, since $\epsilon_\infty$ has a minimum value of unity, the real part of the dielectric factor $\epsilon'$, for the membrane must be considerably larger than unity. Furthermore, since the static dielectric constant, $\epsilon_0$, in turn is much larger than $\epsilon'$ it must be at least ten and may be hundreds. In case the static dielectric constant is as low as ten, the membrane thickness then becomes about 100 A.

In the case of dilute solutions and those liquids and solids which follow equation (2) the dipole rotation is opposed by only a pure viscous resistance and $\alpha = 0$. But as the concentration of a solution such as beta lactoglobulin is increased,[9] or the temperature of a liquid such as glycerin[10] or a solid such as ice is decreased,[11] the value of $\alpha$ increases. This suggests a gradual replacement of the viscous opposition to rotation by intermolecular forces of an elastic nature which becomes almost complete for some solid polymers—and for the *Arbacia* egg membrane. The index $\alpha$ and the intermolecular forces may be expressed as a rather special distribution of relaxation times which varies from a single component for $\alpha = 0$ to an increasingly broad and flat distribution as $\alpha$ approaches unity.[5] If then the cell membrane capacity is of a dielectric nature and this interpretation of liquid and solid dielectrics is correct, the cell membranes are to be thought of as rather solid structures in which the energy of thermal agitation is small or negligible in comparison with the intermolecular potential energy.

*Conductance of the Membrane.*—If equation (1) is a complete and correct description of its electrical characteristics, the cell membrane has no direct-current conductance and consequently is completely impermeable to ions. And indeed it has not been possible to demonstrate a conductance of the membrane with any certainty in the large number of experiments which have been done with small cells of irregular or poorly defined shapes and in media of rather high conductivity. As a result, serious consideration has been given to the possibility that materials can only be transported across the membrane as neutral molecules. However, by the use of long cells such as *Nitella*[12] and various nerve fibers,[13-15] a large cell such as frog egg in pond water[16] or an internal electrode as in *Valonia*[17] it has been possible to measure resistances ranging from 100 ohm-cm.$^2$ to 100,000 ohm-cm.$^2$ for the membranes of a number of cells in a resting state. Since membrane resistances as high as these could not have been detected in the other experiments it is possible to assume that all cell membranes have a similar conductance and ion permeability.[18]

When a cell becomes active,[19] is narcotized,[20] is subjected to changes in the composition of the external medium,[21] is injured or dying or is polarized by current flow[22] this membrane conductance changes. The conductance change correlates closely with the physiological state of the cell and may amount to several orders of magnitude while the change of the capacity, if any, is usually not more than 10 or 20%[19]. On this basis one may picture the capacity as characteristic of an ion-impermeable, inert structure composing usually more than 90% of the effective membrane area with the functional ion permeable aspect relegated to the small remaining area of the membrane.

On the assumption that the membrane has a thickness of 100 A its specific resistance is about $10^{10}$ ohms-cm or about $10^8$ times that of protoplasm and the normal environments of many cells. Although this high resistance may be a result of either a very low concentration or a very low mobility of the ions in the membrane, it is more probably that both of these factors are involved to a considerable extent.

It is to be expected that the resting potential and the conductance of the membrane are both functions of the ion permeability and so should be closely related. At the present time there are only a few cells in which the resting potential has been measured in the absence of current flow and practically none in which the resting potential and the membrane conductance have been measured simultaneously.

The indications in the squid axon[21] are that the membrane resistance decreases for all increases of external potassium concentration, although the membrane potential is but slightly altered until the external potassium is about three times its normal value. Changes of the external calcium seem to be almost without effect on the potential but the resistance becomes very high for a tenfold increase and practically vanishes as the external calcium is reduced toward zero. A more extensive use of internal electrode techniques should produce more satisfactory and complete evidence as to the nature of the relation between these two characteristics of the cell membrane.

*Rectification of the Membrane.*—The current flow through a membrane is proportional to the change of the potential difference for small changes— a few millivolts for the squid axon—and the ratio of current to potential change is quite properly expressed as the conductance discussed above. But several lines of evidence have made it highly probable that the relationship is not so simple for larger changes of potentials as has been verified by direct measurements on nerve fibers[15, 22] and plant cells.[23] As the potential difference was decreased more than a few millivolts the outward current increased more rapidly than it did for small changes, while for the large increases of the potential difference the inward current increased more slowly than for the small changes of potential. This highly non-linear relationship between the current and the potential difference is of the type which characterizes a rectifier. In the squid axon the ratio of the maximum conductance for outward flow to the minimum conductance for inward flow is about a hundred to one.[22] If this effect is ascribed to an ion, such as potassium, which has a higher concentration inside than out, the concentration of this ion should be higher for an outward than for an inward current flow at every point in the membrane. As a result the outward conductance should be higher than the inward which is in qualitative agreement with the experiments.

The problem may be approached quantitatively by means of the equation for the ion concentration, $n$, under the influence of diffusion and electric forces,

$$\rho\frac{\partial n}{\partial t} = -\frac{\partial}{\partial x}\left(neX - kT\frac{\partial n}{\partial x}\right), \tag{4}$$

where $\rho$ is the friction coefficient and $X$ is the electric field. This equation has been applied to the barrier layer rectifier and solved on the assumption of a negligible space charge. Although this assumption does not appear to be justified in the present problem and a more rigorous solution is not available, it has been possible to obtain reasonably good agreement with some of the experimental results on the rectification characteristics and the membrane potential of the squid axon.[24]

*Anomalous Reactance of the Membrane.*—When definitive data were taken on the membrane of squid axon at frequencies below 1 kc. another physical characteristic made its appearance.[25] This is a reactance in addition to the membrane capacity of one microfarad per square centimeter which, unlike the latter, is closely correlated with physiological and chemical factors and may be either capacitative or inductive.[19] The present indications are that it is normally present in the squid axon membrane and somewhat capacitative although it is rather variable. A cursory inspection of the data on other biological systems suggests rather strongly that this anomalous reactance is not a unique feature of the squid axon membrane although indications of an inductive reactance have not been found except in nerve fibers.

The reactance of the squid axon membrane becomes highly capacitative with excess calcium in the external medium and highly inductive as the normal calcium is removed from the medium. An excess of external potassium up to five times the normal value gives rise to an inductive reactance which then disappears for still higher concentrations. The investigations of the effect of membrane current flow on this anomalous reactance have not been satisfactory but the preliminary results indicate that the reactance tends to become inductive for outward current flow and capacitative for an inward flow.

The most attractive explanation of this reactance[3] is that it is a perfectly general consequence of the non-linear characteristic of the membrane. Let us assume that the mechanism which is responsible for the non-linearity of a system cannot respond to instantaneous or sufficiently fast variations. In electrical terms we have $e = r_\infty i$, where $e$ and $i$ are small rapid variations of the potential difference and the current, respectively, and $r_\infty$ is the instantaneous variational resistance. We will now assume that this instantaneous response is modified by the non-linear mechanism to approach the steady-state response, $e = r_0 i$ at a rate $-\beta(e - r_0 i)$, where $r_0$ is the steady state of variational resistance and $\beta$ is the rate constant of the mechanism. The phenomenon is then described by the equation

$$\frac{de}{dt} = r_\infty \frac{di}{dt} - \beta(e - r_0 i).$$

The transient solution for a constant current $i_0$ applied at $t = 0$ is

$$e(t) = [r_0 - (r_0 - r_\infty)e^{-\beta t}]i_0$$

and the steady-state impedance is

$$z(i\omega) = r_\infty + \frac{r_0 - r_0}{1 + i\omega\tau},$$

where $\tau = 1/\beta$. According as $r_0 - r_\infty$ is greater or less than zero both the transient and the steady-state solutions are represented by equivalent circuits containing resistances and a single capacity or inductance, respectively. The values of the reactive elements for one circuit are

$$C = \tau/(r_0 - r_\infty) \text{ and } L = (r_\infty - r_0)\tau.$$

This formal consideration indicates that an anomalous reactance is to be expected in a non-linear system, but for an interpretation of the velocity constant it is necessary to make a detailed analysis of the mechanism involved. Although either a chemical reaction or a transport process may be primarily responsible for this velocity constant we shall use only the simple potassium diffusion model already discussed as an illustration.

In this model, the distribution of ions at equilibrium is such that the outward diffusion is prevented by the inwardly directed electrical field. At the instant when the field is changed by a sudden increase or decrease of the potential difference across the membrane the initial current flow is determined by the ion distribution which was present before the potential was changed. But as time goes on the ion distribution and the current change to approach the steady-state values required by the alteration of the potential. For an increase of potential, the current decreases from its initial value to a lower steady-state value which corresponds to the presence of an equivalent capacity in the membrane. For a decrease of potential, the current rises to a higher value indicating an equivalent inductance. The details of these processes have been calculated from equation (4) and it is found, as can be seen by normalizing the equation, that the order of magnitude of velocity constant is given by

$$\beta = X^2e^2/\rho kT.$$

The available data indicate that $\beta$ is of the order 100 sec.$^{-1}$ for the squid axon. A crude combination of this value and the conductance data leads to an ionic concentration in the membrane $n = 5.10^{17}$ ions/cc. or about 0.001 $N$ and a friction coefficient $\rho = 3.10^{-4}$ dyne sec./cm. which is larger than the value for the potassium ion in water by a factor of about $10^5$.

*Conclusion.*—Although this discussion has been limited to a few of the simpler physical aspects of bioelectric phenomena, it is apparent that no one of the membrane properties mentioned has as yet been adequately explained in terms of physical or chemical principles and that the description of the structure and function of the living cell membrane in elementary terms is still far from complete.

---

[*] Present Address: Naval Medical Research Institute, Bethesda, Md.

[1] Fricke, H., *Phys. Rev.*, **26**, 682–687 (1925).

[2] Cole, K. S., *Tabulae Biologicae (Cellula)*, **19** (2), 24–47 (1941).

[3] Cole, K. S., Four Lectures on Biophysics, Rio (1947).

[4] Cole, K. S., *J. Gen. Physiol.*, **15**, 641–649 (1932).

[5] Cole, K. S., and Cole, R. H., *J. Chem. Phys.*, **9**, 341–351 (1941). Cole, K. S., and Cole, R. H., *Ibid.*, **10**, 98–105 (1942).

[6] Wood, L. A., *Rev. Sci. Inst.*, **4**, 434–439 (1933).

[7] Nolle, A. W., Quarterly Report Acoustics Lab., MIT, April (1947).

[8] Debye, P., Polar Molecules, Chemical Catalog Co., N. Y. (1929).

[9] Shaw, T. M., Jansen, E. F., and Lineweaver, H., *J. Chem. Phys.*, **12**, 439–448 (1944).

[10] Morgan, S. O., *Trans. Electrochem. Soc.*, **65**, 109–116 (1934). Wintsch, H., *Helv. Phys. Acta.*, **5**, 126–144 (1932).

[11] Murphy, E. J., *Trans. Electrochem. Soc.*, **65**, 133–142 (1934).

[12] Blinks, L. R., *J. Gen. Physiol.*, **13**, 495–508 (1930).

[13] Cole, K. S., and Hodgkin, A. L., *Ibid.*, **22**, 671–678 (1939).

[14] Hodgkin, A. L., and Rushton, W. A. H., *Proc. Roy. Soc.*, **B133**, 444–479 (1946).

[15] Hodgkin, A. L., *J. Physiol.*, **106**, 305–318 (1947).

[16] Cole, K. S., and Guttman, R. M., *J. Gen. Physiol.*, **25**, 765–775 (1942).

[17] Blinks, L. R., *Ibid.*, **14**, 139–162 (1930).

[18] Cole, K. S., *Trans. Faraday Soc.*, **33**, 966–972 (1937).

[19] Cole, K. S., and Curtis, H. J., *J. Gen. Physiol.*, **22**, 37–64 (1941). Cole, K. S., and Curtis, H. J., *Ibid.*, **22**, 649–670 (1939).

[20] Guttman, R., *Ibid.*, **22**, 567–591 (1939).

[21] Cole, K. S., and Marmont, G. H., *Fed. Proc.*, **1** (II), 15–16 (1942).

[22] Cole, K. S., and Curtis, H. J., *J. Gen. Physiol.*, **24**, 551–563 (1941). Cole, K. S., and Baker, R. F., *Ibid.*, **24**, 535–549 (1941).

[23] Blinks, L. R., *Ibid.*, **13**, 793–806 (1930).

[24] Goldman, D. E., *Ibid.*, 37–60 (1943).

[25] Cole, K. S., and Baker, R. F., *Ibid.*, **24**, 771–778 (1941).

---

# THE SOURCE OF THE BIOELECTRIC POTENTIALS IN LARGE PLANT CELLS

BY L. R. BLINKS

HOPKINS MARINE STATION OF STANFORD UNIVERSITY

Communicated by D. A. MacInnes, July 8, 1949

The use of large plant cells for the study of bioelectric potentials was originated by Professor Osterhout, and received much of its advance in his laboratory.[1] The advantages of such multinucleate cells are several: they occur either singly, or easily separable from their neighbors with a minimum of dissection or injury, and they survive well in the laboratory, often for days or even weeks with a fine glass tube making connection with the cell sap. The large and measurable surface allows expression of resistance and capacity in definite terms; the capacity being often as high as several microfarads in a single cell, and resistances as high as one megohm (100,000 ohms per square centimeter of surface). New solutions may be quickly applied over the whole surface, or at definitely separated areas for the study of concentration effects, etc. The cell sap may be analyzed for its constituents, or obtained in sufficient quantity to apply to the exterior (see below); in two genera it may be replaced within the vacuole by perfusion with sea water or other new solutions. Through these means the effect of ionic and other changes has been studied at both the outer and inner (or vacuolar) protoplasmic surfaces, leading to a better understanding of the gradients, and the origin of the bioelectric potential.

*Normal* P.D.—Some nine or ten genera of large cells have now been studied. The results are tabulated in table 1. The majority display a normally positive potential difference (P.D.) across the protoplasm, the outer surface being positive to the external measuring circuit. As in the case of many animal cells *Halicystis*, *Bryopsis*, *Batophora*, *Nitella*, *Chara* and *Hydrodictyon* all display positive potentials of 50 to 80 mv. (*Nitella* sometimes higher). However, *Valonia*, perhaps the most studied genus, displays a low negative P.D. of 8 to 15 mv. in the sea water that is its natural environment. Its close relatives *Ernodesmis* and *Chamaedoris* are likewise negative. However, both these negative and positive P.D.'s are reproducible, not only as to sign, but even as to magnitude.

TABLE 1

POTENTIAL DIFFERENCE (P.D.) ACROSS THE PROTOPLASM OF MULTINUCLEATE PLANT CELLS

The sign is that of the outer solution; when positive, the current tends to flow outward across the protoplasm to the outer solution and to the measuring instrument. Values in millivolts. ( ) signifies few observed values; (?) questionable.

| ORGANISM | P.D. IN NORMAL ENVIRONMENT (SEA WATER OR POND WATER), MV. | | P.D. IN NATURAL OR ARTIFICIAL SAP, MV. | |
|---|---|---|---|---|
| Marine: | | | | |
| *Valonia macrophysa* | 5 to 10 | − | 20 to 80 | − |
| *V. ventricosa* | 10 to 20 | − | 20 to 80 | − |
| *Ernodesmis verticillata* | 5 to 10 (?) | − | | |
| *Chamaedoris annulate* | (5) | − | | |
| *Batophora Oerstedi* | 60 to 80 | + | (5 to 10) | + |
| *Halicystis ovalis* | 75 to 80 | + | 20 to 60 | + |
| *H. Osterhoutii* | 65 to 70 | + | 40 to 60 | + |
| *Bryopsis hypnoides* | (50) | + | (0) | |
| *B. corticulans* | 50 to 60 | + | (0) | |
| Fresh water: | | | | |
| *Nitella flexilis* | 100 to 200 | + | 0 to 20 | − |
| *N. mucronata* | 100 | | | |
| *N. elevata* | 100 to 150 | + | | |
| *Chara coronata* | 75 to 100 | + | | |
| *Hydrodictyon patenaeforme* | 50 to 60 | + | 0 to 10 | − |

*Application of Sap: Protoplasmic Asymmetry.*—Since the vacuolar sap of most of these plants differs from the external solution (especially in its potassium content), it is often supposed that the resulting ionic gradients account for the P.D. If so, the latter should become zero when natural or artificial sap is applied externally. Table 1 shows that this is seldom the case. It is most nearly true in *Nitella* and *Bryopsis*, although even in these genera the P.D. is sometimes reversed to slightly negative values by this treatment. In most other genera the P.D. is only slightly lowered, as in *Halicystis*, or is even *increased*, as in *Valonia*. Obviously, in the apparently symmetrical system:

Sap | Protoplasm | Sap

there must be some residual asymmetry within the protoplasm itself, as originally emphasized by Osterhout, *et al.*[2] The simplest assumption is that the two protoplasmic surfaces differ in their response to identical solutions. The nature of this response will be next considered.

*External Cation Effects.*—The outer surface of the cell has long been known to respond electrically to changes in the ion concentration or composition of the medium. Although various causes have been ascribed to these responses (e.g., phase boundary potentials, adsorption effects, colloidal alterations or effect upon metabolism), the most generally useful interpretation has been based on diffusion potentials through the protoplasmic surface or "plasma membrane," likewise taking into account the differing partition coefficients ($S$) of salts in this surface—presumably a non-aqueous phase, since it is not miscible with the cell sap or with water. These effects are clearly very different from those in water, since KCl displays practically no diffusion potential in water, but has a marked influence on the bioelectric potential of most plant cells, increasing in a regular fashion with concentration. In *Nitella* this is perhaps the most striking,[3] where the P.D. change plotted against log of concentration of KCl gives a straight line over a wide range of concentration (0.001 $M$ to nearly 0.1 $M$). The slope of this line is nearly 58 mv. per tenfold dilution, or close to the limit for a perfect "potassium electrode." If the Nernst equation for diffusion potentials is applied, a relative mobility of $K^+$ ion equal to 73 must often be assumed to account for this slope, if that of $Cl^-$ is taken as 1.0. The relative mobility of $Na^+$ in *Nitella* is ordinarily much smaller, usually between 2.0 and 3.0 when $Cl^-$ is unity; this gives a concentration effect of 20 to 30 mv. per tenfold dilution. In both these cases dilute solutions are more positive to the measuring circuit than are more concentrated ones; consequently the normally positive P.D. of *Nitella* is decreased by higher concentrations of either KCl or NaCl. Other cations such as $Li^+$ and $Cs^+$ resemble $Na^+$; $Rb^+$ is nearly as high as $K^+$; while $NH^+_4$, $Ca^+$, and $Mg^+$ lie between these (see table 2). On going from one salt to another, it is frequently necessary to assume a different solubility or partition coefficient ($S$) to account for the change of potential. With this addition mixtures of two salts may also be calculated with the Henderson equation.[4]

In contrast to *Nitella*, in both of the marine genera most studied (*Valonia* and *Halicystis*), the relative mobility of $Na^+$ must be assumed as *less* than that of $Cl^-$, since dilution of the sea water (largely NaCl) causes the

P.D. to become less positive (more negative in the case of *Valonia*, with a negative P.D. already). This dilution effect is especially reproducible in *Valonia*,[5] where a regular stepwise increase of P.D. occurs on successive two-fold dilutions of sea water (or of the NaCl of sea water) with isotonic glycerol, up to $^1/_8$ the normal concentration. A straight line results when the P.D. change is plotted against log (1/ concentration), the slope being such as to demand a relative mobility of 0.114 to 0.33 for $Na^+$ when $Cl^-$ is unity. Somewhat similar values result in *Halicystis* for $Na^+$, though the cells tolerate less dilution, and the P.D. changes are transient, with the time course exhibiting an extremum or cusp.

TABLE 2

RELATIVE MOBILITIES OF IONS IN OUTER PROTOPLASMIC SURFACE OF SEVERAL PLANT CELLS

(Calculated with the Nernst and Henderson equations on the assumption that the salt effects are due to differing relative mobilities in the surface.) Mobility of chloride ion assumed as unity. The figures enclosed in parentheses represent values obtained under special conditions.

| ION | *Valonia macrophysa* | *Halicystis* (*Osterhoutii* or *ovalis*) | *Nitella flexilis* |
|---|---|---|---|
| Cations: | | | |
| K | 18 to 20 | 16 to 18 | 73 (8.7) (2.2) |
| Na | 0.2 to 0.11 | 0.2 to 0.3 | 2.2 (8.7) |
| Li | 0.2 | 0.2 to 0.3 | 2.04 |
| Cs | 0.2 | 0.2 to 0.3 | 1.72 |
| Rb | 18 | 16 (0.3) | 73 (?) (8.76) |
| $NH_4$ | (37) | (16?) | 4.05 |
| Ca | | | 7.5 |
| Mg | | 1.9 | 20.7 |
| Anions | | | |
| Cl | 1.0 | 1.0 | 1.0 (assumed as unity) |
| Br | | (0.5) | |
| $NO_3$ | (0.2?) | 0.1 | |
| Formate, acetate, etc. | (0.2?) | 0.1 | |

Conversely, in both marine genera, $K^+$ appears to have a higher mobility than $Cl^-$; an *increase* of the originally low KCl content of sea water (doubling, quadrupling, etc.) tending to make the P.D. become less positive (or more negative in *Valonia*). But instead of the approximate independence of the time in the regular stepwise effects just mentioned, the time course is strongly cusped, with a prompt decline and a slow recovery. Damon[6] has interpreted this time course as due to the initial diffusion potential of KCl into the protoplasmic surface, followed by the rather rapid entrance of KCl into and across that surface, with consequent abolition of the diffusion gradient, and loss of the diffusion potential. The same transient time course has been found with KCl increase on *Halicystis* by the author,[7] who now prefers to interpret the cusp as due to a "desensitization" of the surface, which ceases to discriminate between $Na^+$ and $K^+$ after 5 to 10 minutes' exposure to the latter. (Temperature influences this,[7] as does pH.[8]) In either case, by taking the maximum value of the cusp, it is possible to calculate an approximate value for the relative mobility of $K^+$, which is between 16 and 20, when that of $Cl^-$ is again assumed as unity.

Although it may well be true that the high mobility of $K^+$, and its high concentration in the cell sap, accounts for the natural P.D. of *Nitella*, this can hardly be the explanation of the P.D. in *Valonia* and *Halicystis*. In these genera: (*a*) the potassium effects are transient, giving rise to no very permanent potential change; (*b*) the P.D. of *Valonia* is in the wrong direction for the gradient to be operative; (*c*) in *Halicystis* there remains a large positive P.D. when sap is applied externally,[7] when sea water is perfused in the vacuole,[8] or in a species (*H. Osterhoutii*) naturally containing less $K^+$ than the sea water itself.[9]

We must conclude that some other gradient must be responsible, and turn to experiments with other ionic species. The most fruitful investigations have concerned anion substitutions.

*Anion Substitutions in Sea Water.*—The most abundant anion of sea water, next to $Cl^-$, is $SO_4^{--}$, which reaches a concentration of about 0.028 *M* whereas $Cl^-$ averages about 0.550 *M*. In both *Valonia* and *Halicystis* sulfate is absent from the cell sap, but its gradient to sea water is not very significant; it may be doubled, or omitted from an artificial sea water, with little effect upon the P.D., showing that it is not important *as such* in the normal environment. However, when the sulfate concentration is considerably increased, the P.D. of *Halicystis* is depressed, and a complete substitution of $SO_4$ for Cl in an artificial "sulfate sea water" produces a very striking effect:[10] the normally positive P.D. of some 80 mv. is reversed to a negative one of 20 to 40 mv.—a total change of over 100 mv. This is not transient, but is well maintained at this negative level for an hour or more. If such a pure sulfate sea water is now diluted with isotonic glycerol, the P.D. becomes less negative, indicating that the $SO_4^{--}$ ion has a lower mobility than has $Na^+$ (which in turn is lower than that of $Cl^-$). If the mobility of $Cl^-$ is again assumed to be unity, that of $Na^+$ is 0.3 and that of $SO_4^{--}$ between 0.1 and 0.2.

Mixtures of sulfate and chloride sea waters give intermediate values, which are very stable, and roughly predictable from the Henderson equation for mixed electrolytes. As might be also predicted, it is possible to make an "indifferent" mixture of these two anions which gives practically no change on dilution with isotonic glycerol.

The behavior of sulfate is paralleled by most of the other anions tolerated by the cells (except bromide, which resembles chloride, though it is somewhat less mobile). Thus nitrate, formate, acetate, propionate, lactate, pyruvate, fumarate, succinate and malate all behaved essentially like sulfate; so did a "sea water" made of the amino-acid salt, sodium glutamate. (Oxalate and citrate were toxic due to their effect on the necessary calcium.) Nitrate and formate were most studied, being most constant and least toxic. Both these anions, when substituted for chloride, maintained the P.D. at a very constant negative value of about 40 mv. Mixtures with normal sea water, containing 0.1 $M$ $NO_3$, 0.2 $M$, 0.3 $M$, etc., maintained predictable intermediate levels. A mixture of 4 or 5 parts $NO_3$ to one of Cl gave an indifferent solution showing little or no change on dilution with glycerol, the negative influence of the NaCl dilution being balanced by the positive one of $NaNO_3$ dilution. Dilution of pure nitrate sea water showed the relative mobility of $NO_3^-$ ion to be about 0.1 when that of $Na^+$ is 0.3 and that of $Cl^-$ is unity. Since the mobility of $K^+$ is 16 or more, we have the remarkable alternating sequence: $K > Cl > Na > NO_3$, etc. Clearly no membrane with charged pores could give rise to this alternation of mobilities for positive and negative ions. That nitrate has a low actual as well as relative mobility in the cell surface is indicated by the increased resistance of cells in contact with it, and by the very persistent P.D. changes produced by it (instead of the cusps produced by the very mobile $K^+$ ion).

*Source of Potential.*—Since a great variety of organic anions behave like nitrate and sulfate in having a very low relative mobility in the outer protoplasmic surface, we may now postulate a possible source of the normal potential. All we require is to have the total organic anions of the cytoplasm approximate 0.5 $M$: in this we may include not only the fatty acids, lactate, pyruvate, etc., actually studied, but a variety of amino acids, and of course proteins, whose anions would doubtless have an even lower mobility in the protoplasmic surface. Analytic data on the very thin layer of cytoplasm in these cells are lacking, but it is hoped to obtain them from the related genus *Codium;* some estimates in other cells make the assumption at least a possible one. But even if such a gradient should set up a large potential across the outer surface, we must know the electrical situation across the vacuolar membrane as well; if the latter behaves like the outer surface, an oppositely directed potential would be set up cancelling the one at the outer. Fortunately, it is possible to learn something about the vacuolar surface in *Halicystis* and *Bryopsis.*

*Vacuolar Perfusion.*—This is accomplished by direct substitution of new solutions in the vacuole, either by two inserted capillary tubes[8] (in *Halicystis*) or by cutting both ends of the cell (in *Bryopsis*). In the latter case the substitution can be accomplished in about a minute; in *Halicystis* this requires 5 to 10 minutes (as shown by tracer dyes). In both cases the replacement of natural sap, fairly rich in KCl, by sea water, has practically no effect upon the P.D. That of *H. ovalis* remains still at 80 mv. positive, despite the absence of any net gradient across the protoplasm, in the apparently symmetrical system:

Sea water | Protoplasm | Sea water.

Clearly again (just as with the external application of sap) a KCl gradient between sap and sea water is not responsible for the potential difference, which must be referred to an internal asymmetry of the protoplasm. The nature of this becomes apparent when anions are changed in the perfusing fluid. When nitrate is substituted for chloride, either in artificial sap or sea water, there is indeed an effect on the P.D., which *increases* from its normal 80 to 100 mv. or more positive. This is to be expected from the low mobility of the $NO_3^-$ ion, compared to $Cl^-$. But the amount of change is only 20 to 30 mv., instead of the 100- to 120-mv. change produced by this same substitution externally. Evidently the inner or vacuolar surface is able to discriminate only $^1/_4$ to $^1/_5$ as well as the outer surface between $Cl^-$ and $NO_3^-$. Formate and several other organic anions again behave like nitrate when perfused. We may therefore conclude that in the scheme suggested below the potential established across the surface ($Y$) by the organic anions of the cytoplasm will be only $^1/_4$ to $^1/_5$ that set up across the outer surface ($X$). Assuming the concentration of these anions ($A$) again at about 0.5 $M$, we might expect the following values:

$$\begin{array}{ccccc} \text{Sea water} & X & \text{Cytoplasm,} & Y & \text{Vacuolar sap} \\ \text{(largely NaCl)} & & \text{NaA, KA ,} & & \text{NaCl, KCl} \\ & \longleftarrow & & \longrightarrow & \\ & \text{100–105 mv.} & & \text{20–25 mv.} & \\ & & \text{Net P.D.} & & \\ & & \longleftarrow\!\!\ll & & \\ & & \text{80 mv.} & & \end{array}$$

The net P.D. is obviously due to the same gradient operating across two different membranes, one of which develops a diffusion potential four or five times as large as the other (and of course in opposite directions).

*Reversal of* P.D.—Independent evidence for the existence and magnitude of two such discrete potentials, giving an algebraically additive total P.D., comes from the effect of agents apparently altering the outer surface independently of the inner. Dilute ammonium salts are most reproducible in this respect: at a threshold of between 0.001 and 0.003 $M$ $NH_4Cl$ in sea water, the P.D. of *Halicystis* suddenly decreases and then reverses in a very characteristic time course.[11] Raising the pH from 8.0 to 9.0 has the same effect in the presence of lower $NH_4Cl$ concentrations. This indicates that the penetration of undissociated ammonia is responsible for the effect, the "all-or-none" nature of which suggests the complete loss of the positive potential at the outer surface ($X$). That at the inner surface ($Y$) inwardly directed, persists at 20 to 30 mv. negative, and is not influenced by further increases of ammonia. It is possible that the entering ammonia, having passed across the lipoid outer surface, combines with the polar groups of the latter, which are oriented toward the cytoplasm. This must saponify, emulsify or otherwise alter the outer surface, so that it loses its power to discriminate between various ions. (Since the resistance rises, it is possible that all ionic mobilities are decreased, to become nearly equal.) This is borne out also by the fact that under ammonia treatment, the outer surface is no longer responsive to anion substitutions: nitrate sea water has no effect when the P.D. is reversed by the weak base. (Various amines, pyridine, aniline and several alkaloids behave like ammonia in producing this effect.) The inner or vacuolar surface, on the other hand, seems to be immune, since the inwardly directed potential persists even under much higher concentrations of $NH_4Cl$, or when the latter is perfused in the vacuole. The effect on the outer surface is quite reversible, normal positive P.D. returning when ammonia is withdrawn (or the pH lowered).

Another agent which causes reversal of P.D. is physiological unbalance of salts bathing the outer surface: if NaCl is substituted for sea water, the P.D. of *Halicystis* reverses temporarily to 20–30 mv. negative, before coming to zero.[12] Addition of calcium causes recovery. Apparently the outer surface ($X$) is first altered by lack of calcium, with loss of its potential, and uncovering of the negative potential at the surface $Y$ before the latter is in turn affected.

*Metabolic Effects: Lack of Oxygen.*—It is well known that many tissues respond to critically low partial pressures of oxygen by a decrease of their bioelectric potential; *Halicystis* is no exception. When the surrounding sea water is bubbled with 1% or less $O_2$ in $N_2$, the P.D. falls from about 80 mv., to about 10 mv. positive.[13] The anion substitution technique again supplies the reason for this. When nitrate is substituted for chloride ($O_2$ being kept low) there is practically no effect on the P.D.; under anaerobic conditions the surface has ceased to discriminate between these two anions. Consequently it must have ceased to give a potential due to their gradient between cytoplasm and sea water. This accounts sufficiently for the fall of potential, even as to time course, since the nitrate effect disappears *pari passu* with the fall of potential under anoxia, and recovers along with the P.D. when oxygen is readmitted. The resistance rises markedly as the P.D. falls. The reasons for this effect of anoxia are interesting, implying some close connection between the cytoplasm and its surface, between the products of metabolism and the electrical properties of the cell membrane. (It is probably not necessary to assume a rapid change of the ionic gradients themselves, but only of the phase which gives electrical expression to them.) This might consist of the loss of some substance constantly produced in aerobic metabolism (see below), or the increase of some anaerobic fermentation product. It is possibly due to an increase of cytoplasmic acidity (e.g., lactic acid), for much the same effects are produced by weak acids which penetrate the protoplasm, such as carbonic or acetic: the P.D. likewise falls, the anion effect disappears and the resistance rises.

*Other Chemical Effects.*—The membrane properties can be affected by a variety of other substances, some in considerable dilution. Thus a number of weak acids and several phenolic substances (phenol, cresol, guaiacol) strongly influence the P.D. of *Valonia*, rendering it highly positive (like the normal P.D of *Halicystis*). Osterhout[14] found that this positivity was accompanied by a marked change of ionic mobilities in the protoplasmic surface, Na and K reversing their usual positons: thus if the mobility of $Cl^-$ was assumed to be 1.0, that of $K^+$ became 0.36 (instead of 20), and that of $Na^+$ 4.5 instead of 0.3 (see table 2). In *Halicystis* exposed to such agents, Na was also raised above Cl, although K was not affected.[15] The direct-current resistance of both *Valonia* and *Halicystis* rises under these treatments, and marked polarization curves appear (likewise under anoxia) indicating that a change of absolute, as well as of relative, ionic mobilities has occurred in the protoplasmic surfaces. Benzene, nitrobenzene, dinotrophenol, hydroquinone, etc., have somewhat similar effects, possibly through alteration of the pathways of normal respiration, and the accumulation of certain intermediate metabolites. Cyanide and urethane lower the P.D. of *Halicystis* somewhat, but not as drastically as several of the treatments mentioned. Indeed, the actual *rate* of aerobic metabolism does not seem to be directly correlated with the P.D. Thus low temperatures, though decreasing respiration (with a respiration quotient for a 10° interval, $Q_{10}$, of 2 or more), scarcely lower the P.D. which has in *Halicystis* a $Q_{10}$ of 1.2, i.e., a 20% decrease for a 10° interval, or about that of diffusion. In fact, the P.D. of *Valonia* may even become higher at low temperatures, probably because the entrance of K is slowed down, with abolition of the cusps of the KCl curves (see above) allowing the KCl of sea water to be more effective.[7] Nor does increased temperature affect P.D. as much as it does respiration. A direct dependence of P.D. on respiration is not indicated.

*Effects of Oxidants and Reductants.*—Some theories[16] have emphasized a direct connection between respiration and bioelectric potentials, via oxidation-reduction potentials as such. This implies that a continuous phase of the cell itself, not appreciably short circuited, is interposed between two regions of different redox potential, and can act as an electron acceptor and donor, like a gold or platinum electrode, to make such potentials manifest in a measuring circuit. The only such continuous phases which are known are the protoplasmic surfaces. Those accessible to experimental treatment, in *Valonia* and *Nitella* do not so function.[17] To test this experimentally, solutions bathing *Nitella* were poised with a variety of oxidants or reductants ranging from ferricyanide ($E_h = +0.454$ volt) to reduced safranin

(−0.288 volt) a total range of 0.740 volt in redox potential. Intermediate systems included ascorbate and dehydro-ascorbate; di-chloro-phenol-indophenol; and indigo-disulfonate. When such systems were oxidized or reduced, with individual redox changes of 200 to 300 mv. none of them showed any appreciable effect upon the P.D. of *Nitella*. (The cells responded normally, however, to different concentrations of NaCl or KCl.) *Valonia* has proved likewise indifferent, except when very powerful oxidants such as permanganate or peroxide were employed; these do slowly alter the P.D., probably through direct attack on the membrane substance, which then shows altered ionic responses. In view of the large response given by various concentrations of neutral salts, both inorganic and organic, it seems unnecessary to invoke redox potentials themselves as the cause of bioelectric potentials. This is in agreement with the chemical difficulties met in considering lipoid or other membrane constituents as electron donors or acceptors.

*Summary.*—Bioelectric potentials may be considered as due to gradients of inorganic and organic ions between protoplasm and sap or environment, coupled with protoplasmic surfaces across which the ions have markedly different mobilities, giving rise to diffusion potentials. The outer and vacuolar surfaces have different properties, giving rise to electrical asymmetries. The high mobility of $K^+$ may be responsible for the P.D. in some cases (e.g., *Nitella*), but the very low mobility of organic anions seems important in others (e.g., *Halicystis*). The properties of the surface are under control, anoxia, ammonia and phenolic substances being especially effective in altering the apparent mobilities of ions through them. Oxidation-reduction potentials do not seem to be directly involved.

[1] For reviews of earlier work from his laboratory, see Osterhout, W. J. V., *Bull. Natl. Research Council*, No. 69, 170–228 (1929); *Biol. Rev.*, **6**, 369–411 (1931); *Ergeb. Physiol. expll. Path.*, **35**, 967–1021 (1933). References to earlier work are given therein.

[2] Osterhout, W. J. V., Damon, E. B., and Jacques, A. G., *J. Gen. Physiol.*, **11**, 193–205 (1927).

[3] Hill, S. E., and Osterhout, W. J. V., *Ibid.*, **21**, 541–556 (1938).

[4] Hill, S. E., and Osterhout, W. J. V., Proc. Natl. Acad. Sci., **24**, 312–315 (1938).

[5] Damon, E. B., and Osterhout, W. J. V., *J. Gen. Physiol.*, **13**, 445–457 (1930).

[6] Damon, E. B., *Ibid.*, **16**, 375–395 (1932); **21**, 383–409 (1938).

[7] Blinks, L. R., *Ibid.*, **16**, 147–156 (1932); **25**, 905–916 (1942).

[8] Blinks, L. R., *Ibid.*, **18**, 409–420 (1935).

[9] Blinks, L. R., and Jacques, A. G., *Ibid.*, **13**, 733–737 (1930).

[10] Blinks, L. R., *Cold Spring Harbor Symposia Quant. Biol.*, **8**, 204–215 (1940).

[11] Blinks, L. R., *J. Gen. Physiol.*, **17**, 109–128 (1933).

[12] Blinks, L. R., Rhodes, R. D., and McCallum, G. A., Proc. Natl. Acad. Sci., **21**, 123–125 (1935).

[13] Blinks, L. R., Darsie, M. L., and Skow, R. K., *J. Gen. Physiol.*, **22**, 255–279 (1938).

[14] Osterhout, W. J. V., *Ibid.*, **20**, 13–43 (1936).

[15] Osterhout, W. J. V., *Ibid.*, **21**, 707–720 (1938).

[16] Lund, E. J., *J. Exp. Zool.*, **51**, 265–337 (1928).

[17] Blinks, L. R., and Pickett, M. J., *J. Gen. Physiol.*, **24**, 33–44 (1940).

# *QUINOLINIC ACID ACCUMULATION IN THE CONVERSION OF 3-HYDROXYANTHRANILIC ACID TO NIACIN IN NEUROSPORA**

BY DAVID M. BONNER AND CHARLES YANOFSKY

OSBORN BOTANICAL LABORATORY, YALE UNIVERSITY, NEW HAVEN, CONN.

Communicated by E. W. Sinnott, August 15, 1949

3-Hydroxyanthranilic acid has been shown to serve as a natural precursor of niacin in Neurospora.[1, 2] Little information, however, has been obtained concerning the mechanism whereby the aromatic ring of 3-hydroxyanthranilic acid is converted to the pyridine ring of niacin. The present paper will deal with this latter problem.

Singal, *et al.*,[3] have reported that the administration of tryptophan to rats results in the urinary excretion of a compound possessing niacin activity for *Lactobacillus arabinosus* after autoclaving in acid solution. This compound has recently been isolated and identified by Henderson[4] as quinolinic acid (2,3-dicarboxypyridine). These data suggest that quinolinic acid plays a rôle in the biosynthesis of niacin. Quinolinic acid had been tested earlier on certain mutant strains of Neurospora requiring niacin for growth, and found to be inactive.[1,2] However, it seemed desirable to determine whether this compound might not be accumulated by niacin-requiring mutant strains, and have escaped attention because of its inactivity. One mutant strain has been found which accumulates this compound as a result of a genetic block.

*Experimental.*—The various mutant strains of Neuorspora used in this investigation, together with the compounds known to support their growth are listed in table 1. The basal medium used throughout is the customary Neurospora minimal,[5] and growth tests were made for 72 hours at 25°C. in 125-cc. Erlenmeyer flasks containing 20 cc. of medium, unless otherwise stated. The *L. arabinosus* assay for niacin is that described by the Association of Vitamin Chemists.[6]

TABLE 1

GROWTH OF VARIOUS STRAINS OF NEUROSPORA

| STRAIN NO. | GROWTH ON MINIMAL MEDIUM PLUS | | | | | | |
|---|---|---|---|---|---|---|---|
| | NO ADDITION | NIACIN | QUINOLINIC ACID | 3-OH-ANTHRANILIC ACID | KYNURENINE | TRYPTOPHAN | ANTHRANILIC ACID |
| 3416 | − | + | − | − | − | − | − |
| 4540 | − | + | − | − | − | − | − |
| Y31881 | − | + | − | + | − | − | − |
| 39401 | − | + | − | + | + | + | − |
| 75001 | − | − | − | − | + | + | + |
| 10575 | − | − | − | − | − | + | − |
| Wild type | + | + | + | + | + | + | + |

Of the mutant strains which require niacin for growth, only strains 4540 and 3416 are known for which some known substance cannot replace the niacin requirement. These two strains are both genetically and biochemically different. Crosses between 3416 and 4540 give wild-type progeny,[7] and strain 4540 has been shown to accumulate 3-hydroxyanthranilic acid,[1] while strain 3416 does not. Of the various niacin-requiring mutant strains, therefore, only strain 3416 might reasonably be suspected of quinolinic acid accumulation since the other strains are genetically blocked prior to the utilization of 3-hydroxynathranilic acid.

Extensive tests of strain 3416 culture filtrates failed to show significant accumulation of an intermediate which would support the growth of any other niacin-requiring strains. Such culture filtrates, however, when autoclaved for 15 minutes at 15-lb. pressure in 1 *N* HCl had marked niacin activity by Lactobacillus assay (see table 2) whereas similar filtrates not so treated had little niacin activity. For the unequivocal demonstration of the presence of quinolinic acid in culture filtrates of strain 3416, sufficient culture filtrate was prepared to permit the isolation of this dicarboxy acid.

TABLE 2

QUINOLINIC ACID ACCUMULATION BY STRAIN 3416

| 3416 CULTURE FILTRATE TESTED IN ML. | γ NIACIN | | γ QUINOLINIC ACID CONVERTED TO NIACIN PER ML. FILTRATE | γ QUINOLINIC ACID PER ML. FILTRATE |
|---|---|---|---|---|
| | BEFORE ACID AUTOCLAVING | AFTER ACID AUTOCLAVING | | |
| 0.03 | 0.02 | 0.3 | 9.3 | 78 |
| 0.003 | 0 | 0.03 | 10 | 83 |
| Quinolinic acid 3.3 γ | 0 | 0.4<br>(12% conversion) | .. | .. |

The method of isolation used was a modification of the method outlined by Henderson.[4] Nineteen liters of basal medium contained in a 5-gal. Pyrex bottle, supplemented with 10 mg. of niacin, were inoculated with strain 3416 and the strain grown with forced aeration for seven days at 25°C. The mycelium was removed by straining through cheesecloth, and the culture filtrate concentrated to ca. 2 liters. This concentrate was filtered and acidified to pH1, and shaken with 100 g. of Norit A. After filtration the norit was treated with dilute ammonium hydroxide. This eluate was evaporated to dryness, and the residue treated with about 30 cc. of methanol slightly acidified with HCl. The entire methanol-soluble fraction was then put on a single sheet of Schleicher and Schuell paper No. 470-A (0.025 inch thick) by the method of Yanofsky, *et al.*,[8] and developed as an ascending chromatogram using butanol saturated with water containing 8% formic acid as the developing solvent. Pure quinolinic acid was run with the same solvent to determine its Rf value. The quinolinic acid band was cut from the dried sheet, eluted with dilute $NH_4OH$ in a Waring blendor and the resulting eluate evaporated to dryness. Quinolinic acid was then crystallized from slightly acidified dilute ethanol. Recrystallization gave a material which began to sublime at ca. 165°, and melted with decomposition and gas evolution at 188–190°. When this material was mixed with authentic quinolinic acid no depression in the melting point was observed. Figure 1 gives a comparison of the absorption spectra of the isolated material and authentic quinolinic acid in 0.1 *N* HCl. The crystalline material was inactive for growth of *L. arabinosus*, but showed conversion to niacin after autoclaving in acid solution. The isolated material, therefore, is identified as 2,3-dicarboxypyridine.

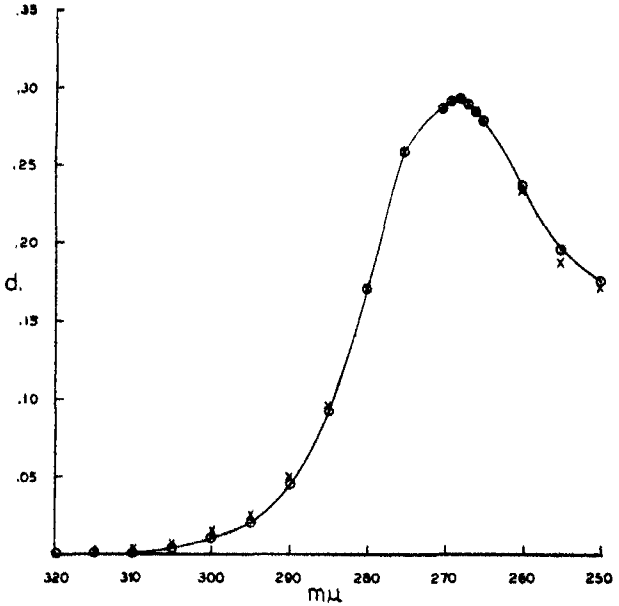


FIGURE 1

Comparison of absorption spectra of quinolinic acid isolated from strain 3416 culture filtrates, and authentic material. ⊙—⊙ quinolinic acid. *x*—*x* isolated material.

In evaluating the significance of these observations, a second point must be clarified. The accumulation of quinolinic acid appears to be associated with niacin synthesis in rats and dogs. Is this accumulation characteristic of the synthesis of niacin by Neurospora, or is it a unique property of strain 3416? A variety of strains were, therefore, tested for quinolinic acid accumulation as shown in table 3. Quinolinic acid accumulation is clearly associated only with strain 3416, and may therefore be considered as peculiar to that particular genetic block.

*Discussion.*—The demonstration by Singal, *et al.*,[3] and by Henderson[4] that quinolinic acid results from the administration of large amounts of tryptophan to rats suggests that quinolinic acid is involved in the conversion of tryptophan to niacin. Henderson[4] further showed that quinolinic acid resulted from the intraperitoneal injection of 3-hydroxyanthranilic acid, suggesting that quinolinic acid serves as an intermediate in the conversion of 3-hydroxyanthranilic acid to niacin.

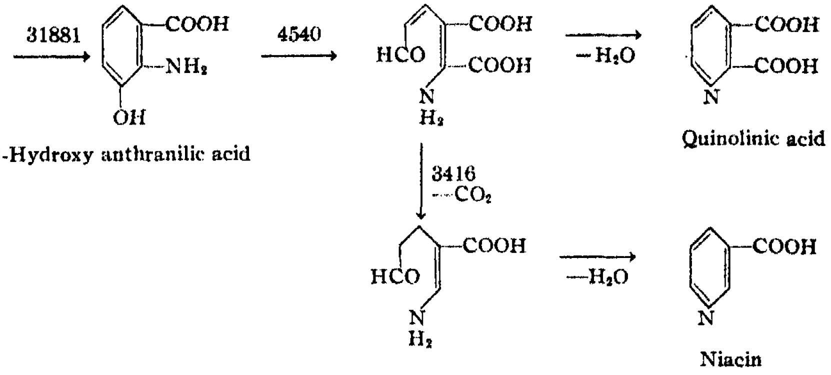


FIGURE 2

Proposed scheme for the conversion of 3-hydroxyanthranilic acid to niacin and quinolinic acid.

TABLE 3

QUINOLINIC ACID ACCUMULATION BY STRAINS OF NEUROSPORA*

| STRAIN NO. | γ NIACIN PER CC. FILTRATE<br>BEFORE ACID AUTOCLAVING | AFTER ACID AUTOCLAVING |
|---|---|---|
| 3416 | 0 | 9.3 |
| 4540 | 0.4 | 0.4 |
| Y31881 | 0.4 | 0.4 |
| 39401 | 0.3 | 0.9 |
| 75001 | 0.4 | 0.4 |
| 10575 | 0.0 | 0.0 |
| Wild type | 0.0 | 0.0 |

* Not corrected for per cent conversion.

The observations in Neurospora are in substantial agreement with those made on the rat, with the exception that quinolinic acid is inactive even for those mutants which are known to be capable of converting 3-hydroxy anthranilic acid to niacin. This inactivity of quinolinic acid in turn suggests a mechanism of conversion shown in figure 2. This scheme suggests that 3-hydroxyanthranilic acid serves as the precursor both of niacin and quinolinic acid. That 3-hydroxyanthranilic acid serves as a precursor of niacin has been previously demonstrated.[1,2] That 3-hydroxy anthranilic acid also serves as a precursor of quinolinic acid may be argued from the fact that quinolinic acid is accumulated by a strain which cannot either utilize 3-hydroxyanthranilic acid (in place of niacin) or accumulate it. All other niacin-requiring mutant strains of Neurospora possess one or the other of these two abilities. Quinolinic acid accumulation must, therefore, result from a block interposed in the conversion of 3-hydroxyanthranilic acid to niacin. Further evidence in this connection is the fact that 3-hydroxyanthranilic acid increases quinolinic acid accumulation by strain 3416, and that in preliminary experiments with washed resting 3416 mycelium, quinolinic acid results from the addition of 3-hydroxyanthranilic acid. The scheme given in figure 2 further suggests cleavage in the 3,4 position with the formation of a dicarboxy aliphatic amino aldehyde. No evidence for this step is available other than the general logic of the entire scheme and the fact that 3416 does not accumulate an active substance but does accumulate quinolinic acid. The proposed intermediate might be expected to undergo ring closure very readily. In view of the inactivity of quinolinic acid, however, ring closure must normally be preceded by decarboxylation in the formation of niacin. If decarboxylation cannot occur, as the result of a genetic block, ring closure would result in quinolinic acid formation. It should be pointed out that the validity of such a scheme rests upon the demonstration of the proposed intermediates. The proposed scheme, however, is in agreement with present data and gives, perhaps, a clearer insight into the general problem of the biological formation of pyridine ring compounds.

*Summary.*—Quinolinic acid has been isolated from culture filtrates of a mutant strain of Neurospora requiring niacin for growth. Quinolinic acid is inactive in supporting growth of niacin-requiring strains. The observation that an inactive compound, quinolinic acid, is accumulated as the result of a genetic block, suggests a mechanism for the conversion of the benzene ring of 3-hydroxyanthranilic acid to the pyridine ring of niacin.

Quinolinic acid has been found to have slight activity for strains 4540, 39401 and 31881. The activity of this acid, after correction for per cent dissociation ($pk_1 = 3.0$, $pk_2 = 5.0$) is at most one-fifth that of niacin. In view of this low activity, and the fact that the precursor of quinolinic acid, 3-hydroxyanthranilic acid, is as active as niacin, the authors feel that the scheme presented in figure 2 is still the most logical.

---

* These investigations were supported in part by the Williams-Waterman Fund for the combat of dietary diseases, and in part by the Rockefeller Foundation.

[1] Bonner, D., PROC. NATL. ACAD. SCI., **34**, 5 (1948).

[2] Mitchell, H. K., and Nye, J. C., *Ibid.*, **34**, 1 (1948).

[3] Singal, S. A., Briggs, A. P., and Hankes, L. V., *Proc. Soc. Exptl. Biol. Med.*, **70**, 26 (1949).

[4] Henderson, L. M., *J. Biol. Chem.*, **178**, 1005 (1949).

[5] Beadle, G. W., and Tatum, E. L., *Am. J. Bot.*, **32**, 678 (1945).

[6] The Association of Vitamin Chemists, Methods of Vitamin Assay. Interscience Press, New York (1947).

[7] Bonner, D., and Beadle, G. W., *Arch. Biochem.*, **11**, 319 (1946).

[8] Yanofsky, C., Wasserman, E., and Bonner, D., *Science*, in press.

---

# *THE RÔLE OF ORGANIC PEROXIDES IN THE INDUCTION OF MUTATIONS**

By Frank H. Dickey,† George H. Cleland and Carol Lotz‡

Kerckhoff Laboratories of Biology, California Institute of Technology, Pasadena, California

Communicated by G. W. Beadle, August 15, 1949

The discovery by Wyss, Stone, and Clark[1] that bacteria grown on a substrate recently exposed to ultra-violet light are subject to high mutation rates shows clearly that some meta-stable chemical substance, probably of no great complexity, is an intermediate in at least a part of the mutagenic action of ultra-violet light. It was supposed that hydrogen peroxide might be responsible for these results, but subsequent work has shown that this cannot be the whole explanation.[2] However, organic peroxides are known to be formed by the action of ultra-violet light on many compounds and such peroxides might very well be the intermediate agents producing the substrate irradiation effect.

The process by which organic compounds, especially ethers, olefins and aldehydes, form peroxides simply on contact with molecular oxygen is not wholly understood. In many simple cases, however, a chain reaction of the sort pictured below appears to be involved.

$$\begin{aligned} R{-}H &\rightarrow R{-} + H \\ R{-} + O_2 &\rightarrow R{-}OO{-} \\ R{-}OO{-} + R{-}H &\rightarrow R{-}OO{-}H + R{-} \end{aligned}$$

The peroxide-forming process is catalyzed by ultra-violet light[3] which, presumably, supplies the energy for breaking a carbon-hydrogen bond in the first step.

The hypothesis that organic peroxides play an essential rôle in the mutagenic action of ultra-violet light has been under investigation in this laboratory for some time. One result of this work, and the subject of the present writing, is the discovery that many simple organic peroxides increase mutation rates.

*Testing Procedure.*—The organism used for detecting and comparing mutagenic agents has been an adenineless, colonial strain of *Neurospora crassa*, the double mutant 70007-38701. The adenineless character in this strain is subject to a low spontaneous rate of reversion to adenine independence. These occurrences are probably back mutations at the adenineless locus, but this has not been demonstrated in the present study. The colonial character, introduced into the strain to permit plate counts, appears to be quite stable. Following essentially a method described by Giles and Lederberg[4] and by Westergaard and Mitchell,[5] conidial suspensions were exposed to various peroxides and the effects calculated from the fraction of these spores that gave rise to colonies on adenine-free medium. Since as many as $2 \times 10^8$ spores may be conveniently treated in a single experiment, this method is capable of detecting very weak mutagenic activity.

In a typical experiment a thoroughly mixed suspension of two-day old spores was divided into four portions, centrifuged and decanted. Two portions were retained as controls and the other two were re-suspended in an aqueous solution of the peroxide and allowed to stand at room temperature for 30 minutes. After washing to remove the treating solution, the concentration of conidia in each centrifuge tube was determined with a hemocytometer, and dilution platings on adenine-supplemented medium were made from each sample to indicate conidial viability and percentage mortality. Finally, the suspensions were spread on a series of adenine-free plates and incubated at 25°C. Mutants formed distinct colonies that could be counted during the third or fourth day after treatment.

Mutation rates were calculated by dividing the number of mutants counted by the number of spores plated. By subtracting from the mutation rate shown by spores subjected to a particular treatment the spontaneous rate shown by untreated spores from the same spore batch there was obtained the quantity termed the "induced mutation rate." It is not practical to determine an average value for the fraction of untreated spores that produce adenine-independent colonies and to use such a fixed figure for correcting the observed mutation rates. Although the spontaneous rate is lower than $0.8 \times 10^{-7}$ in more than 75% of the spore batches, occasionally this rate is very high (e.g., $13 \times 10^{-7}$), presumably as a result of an early mutation in the culture from which the spores were obtained. The same problem has been described by Delbrück[6] in connection with mutations in bacteria.

It appears that the largest numbers of mutants are obtained with treatments that kill 60 to 80% of the spores. Mutation rates based on the numbers of spores surviving the treatment might increase up to very high mortalities but in the present investigation more consistent values have been obtained for mutation rates based on the number of spores treated. Accordingly, treatments producing the greatest actual numbers of viable mutants have been sought.

A danger in working with very high mortalities is the occasional appearance of numbers of adenineless colonies, apparently sustained by adenine released to the medium from dead spores. Such false mutants can usually be distinguished by their frail appearance and their failure to develop beyond an early stage, and always by their inability to grow when transferred to minimal medium.

*Demonstration of the Mutagenic Action of Peroxides.*—Table 1 illustrates the mutagenic action of *tert*-butyl hydroperoxide at various concentrations. The data are taken from typical experiments selected from an extensive study of this material. The table is intended to show the effect of concentration of the agent and to give a rough idea of the reproducibility of the results at a particular concentration. The first objective of this work has been to show that peroxides induce mutations and the values obtained for induced mutation rates are only approximate. Variations in the conditions of treatment, especially temperature, may account in part for the irregularities in the results.

TABLE 1

MUTAGENIC ACTION OF *tert*-BUTYL HYDROPEROXIDE

(Treatment: 30 minutes exposure to an aqueous solution of *tert*-butyl hydroperoxide at the indicated concentration)

| CONCENTRATION, MOLES PER LITER | SPORES TREATED, MILLIONS | MUTANTS | MUTATION | RATES × $10^7$ | MORTALITY, % |
|---|---|---|---|---|---|
| | | | OBSERVED | INDUCED | |
| 0.004 | 48 | 6 | 1.2 | 0.5 | 5 |
| 0.010 | 96 | 51 | 5.3 | 2.5 | 42 |
| 0.089 | 70 | 106 | 15.1 | 15.0 | 29 |
| 0.089 | 72 | 140 | 19.3 | 19.2 | 38 |
| 0.089 | 40 | 51 | 12.8 | 10.5 | 80 |
| 0.089 | 29 | 21 | 7.2 | 7.1 | 91 |
| 0.11 | 32 | 6 | 1.9 | 1.8 | 85 |
| 0.27 | 38 | 0 | 0.0 | 0.0 | 100 |

In Table 2 averages of the results of many experiments with six different peroxides are presented. The concentrations of the respective agents, shown in this table, are the ones which have given the highest mutation rates. For comparison the effects of four established mutagenic agents have been included and also the results of experiments with four mildly toxic substances which had no definite effect on mutation rates. It might be noted that two of the last named group, phenol and formaldehyde, have been reported to have a weak mutagenic action in experiments with Drosophila.[7, 8]

The active principles in the mixtures of hydrogen peroxide with formaldehyde and acetone may be, respectively, $HO{-}CH_2{-}OO{-}CH_2{-}OH$ and $HO{-}C(CH_3)_2{-}OO{-}H$. Hydroxymethyl *tert*-butyl peroxide, derived similarly from the interaction of *tert*-butyl hydroperoxide and formaldehyde, is a well-defined compound.[9] The peroxide derived from diisopropyl ether was obtained by extracting old samples of the ether with water and freeing the extract of volatile material by passing an air stream through it.

TABLE 2

MUTAGENIC ACTION OF VARIOUS AGENTS

(Treatment: 30 minutes exposure to indicated aqueous solutions)

| AGENT | CONCENTRATION, MOLES PER LITER | TOTAL SPORES, MILLIONS | INDUCED MUTATION RATE × $10^7$ | MORTALITY, % |
|---|---|---|---|---|
| Hydrogen peroxide | 0.21 | 1200 | 1.8 | 40 |
| *tert*-Butyl hydroperoxide | 0.089 | 211 | 14.5 | 50 |
| Hydroxymethyl *tert*-butyl peroxide | 0.089 | 270 | 16.3 | 68 |
| Peroxide derived from diisopropyl ether | 0.15 | 146 | 23.0 | 61 |
| Hydrogen peroxide and formaldehyde | 0.022<br>(0.033 in $CH_2O$) | 464 | 8.8 | 77 |
| Hydrogen peroxide and acetone | 0.21<br>(1.36 in $(CH_3)_2CO$) | 96 | 10.4 | 32 |
| X-rays | * | 61 | 15.8 | 37 |
| Ultra-violet light | † | 418 | 46.0 | 32 |
| *Bis*(β-chloroethyl)sulfide | 0.0002 | 103 | 4.0 | 70 |
| *Bis*(β-chloroethyl)methylamine | 0.004 | 279 | 3.6 | 85 |
| Phenol | 0.077 | 465 | −0.4 | 87 |
| Formaldehyde | 0.0244 | 251 | 0.5 | 73 |
| Potassium permanganate | 0.00052 | 383 | 0.0 | 35 |
| *tert*-Butyl alcohol | 1.75 | 275 | −0.3 | 84 |

* Approximately 30,000 r units.

† Exposure of an aqueous suspension with agitation for 75 seconds to 30-watt General Electric germicidal lamp at a distance of 10 cm. The incident ultra-violet energy, nearly all in the vicinity of 2537 A. U., was 6000 ergs/mm.$^2$/min.

The peroxide concentration in the resulting solution was determined iodometrically.

Mustard gas, *bis*(β-chloroethyl)sulfide, and the nitrogen mustard, *bis*(β-chloroethyl)methylamine, were used in borate (pH 8) and acetate (pH 5) buffers, respectively. The solution of the sulfur compound contained, in addition, 1.36 moles per liter of acetone to increase its solubility.

In presenting this demonstration of the mutagenic action of peroxides it is assumed that the adenine-independent colonies appear as the result of true mutations. The action of the peroxides certainly produces a heritable change since adenine independence of the reverted strains persists through repeated subcultures on an adenine-free medium. It has been the general experience of other workers that non-genetic reversions of Neurospora mutants, which occur spontaneously in some strains, do not persist through subcultures. This, together with the observation that known mutagenic agents (radiations, mustard gas) produce effects similar to those of the peroxides, constitutes presumptive evidence for the view that the action of the peroxides is mutational. It is recognized, however, that the results of genetic crosses will constitute the only proof that gene mutations are involved. A genetic study directed toward this end is now in progress and will be reported in a later communication.

There can be little doubt that the peroxide treatment *induces* the presumed mutations. Selection cannot account for the results since the spores are not growing during the treatment and since it is an increase in the actual number of adenine-independent colonies that is observed and not merely a higher ratio of the number of such colonies to the number of surviving spores.

Conceivably, certain nuclei (the conidiospores are multinucleate) possess a latent capacity for adenine synthesis that is only made manifest by the peroxide treatment. However, there is no precedent for such a phenomenon. This question as well as the question of the mutational nature of the peroxide-induced changes will be resolved by the current genetic studies mentioned above.

Finally, it should be noted that the experimental method used here is of a rather special sort. In all probability, only the back mutation of a single gene is involved. It will be important to confirm these findings by different techniques and in different organisms.

*Conclusions.*—The foregoing evidence that organic peroxides are capable of producing mutations lends substantial support to the idea that ultraviolet light produces mutations by forming such compounds in irradiated media and in cells. A demonstration that this is the mechanism of the substrate-irradiation effect may eventually be obtained by correlating the mutagenic action of irradiated material with peroxide content.

The facts that so few mutation-inducing agents are known and that the effects of those that are known are in large measure similar suggest that they have a common mode of action. Accordingly, the possibility that peroxides are in some way involved in the action of x-rays and of the mustards is currently under investigation. Possibly compounds of the mustard group or certain products of their interaction with organic compounds are especially liable to peroxide formation.

By determining just what feature of the chemistry of peroxides is responsible for their mutagenic action one might hope to shed light on the nature of the mutation process. It seems unlikely that this action is simply related to oxidizing power, since oxidizing agents are common and organic peroxides are not especially effective ones. Of more interest is the characteristic decomposition of peroxides by which free radicals are produced. If this is the essence of peroxide action non-peroxidic free radical sources (e.g., diazomethane) should show similar effects. It should be noted that irradiation of a cell could produce free radicals directly as well as by peroxide formation.

Besides affording a basis for speculation on the nature of the mutation process, the discovery of the mutation-inducing power of organic peroxides substantially increases the number of known mutagenic agents. Organic peroxides of widely varied structure can be prepared. It will be of interest to compare the action of these various agents on different genes and to search for agents having pronounced effects on particular genes.

[*] This investigation was supported in part by a research grant from Merck and Co.

[†] DuPont Predoctoral Fellow in Chemistry 1948–1949.

[‡] The authors wish to thank Professor Norman H. Horowitz for valuable assistance and advice.

[1] Wyss, O., Stone, W. S., and Clark, J. B., *J. Bact.*, **54**, 767–772 (1947).

[2] Wyss, O., Clark, J. B., Haas, F., and Stone, W. S., *J. Bact.*, **56**, 51 (1948).

[3] Milas, N. H., *J. Am. Chem. Soc.*, **53**, 221–233 (1931).

[4] Giles, N. H., Jr., and Lederberg, E. Z., *Am. J. Bot.*, **35**, 150–157 (1948).

[5] In an unpublished investigation this method for detecting chemical mutagens was developed by M. Westergaard and H. K. Mitchell to whom the authors are also indebted for the stock.

[6] Delbrück, M., *Ann. Mo. Bot. Gard.*, **32**, 223–233 (1945).

[7] Hadorn, E., Rosin, S., and Bertani, D., *Proc. 8th Int. Congr. Gen.*, 256–266 (1949).

[8] Rapoport, I. A., *Acad. Sci. U.R.S.S.*, **54**, 65–67 (1946).

[9] Dickey, F. H., Rust, F. F., and Vaughan, W. S., *J. Am. Chem. Soc.*, **71**, 1432–1434 (1949).

---

# *CROSSING-OVER BETWEEN ALLELES AT THE LOZENGE LOCUS IN DROSOPHILA MELANOGASTER*

BY M. M. GREEN AND K. C. GREEN

DEPARTMENT OF ZOOLOGY, UNIVERSITY OF MISSOURI, COLUMBIA, MISSOURI

Communicated by L. J. Stadler, July 22, 1949

It has been reported that females of *Drosophila melanogaster* having one $X$-chromosome containing the lozenge allele glossy ($lz^g$) and the other $X$-chromosome the lozenge allele spectacle ($lz^s$), crossed to either $lz^g$ or $lz^s$ males, occasionally produce individuals wild type ($lz^+$) in appearance.[1,2]

The occurrence of wild-type progeny from this cross is unexpected since all the progeny of both sexes should have the phenotype of either $lz^s$ or $lz^g$. Furthermore, it has been noted that the production of the non-mutant $X$-chromosome has resulted when crossing over has occurred in the vicinity of the lozenge locus. These results suggest that either the $lz^+$ chromosome occurred as a result of unequal crossing over between $lz^g$ and $lz^s$ in the same manner as Bar in *D. melanogaster*,[3] or that $lz^g$ and $lz^s$ are in fact not alleles in the usual sense but represent two closely linked loci, such that an ordinary crossover between them would yield an $lz^+$ chromosome.

In view of the information already obtained, experiments were set up to determine (1) whether the phenomenon of crossing over between $lz^g$ and $lz^s \rightarrow lz^+$ was peculiar to these lozenge alleles, or whether it occurred between other lozenge mutants as well; (2) whether the crossing over involved is equal or unequal; and (3) what is the nature of the complementary crossover $X$-chromosome, presumably bearing two lozenge mutants.

TABLE 1

RECOMBINATION RESULTS FROM ♀♀ HETEROZYGOUS FOR TWO DIFFERENT LOZENGE ALLELES CROSSED TO ♂♂ *In(1)dl 49, v* $lz^g$

| GENOTYPE OF ♀♀ | TOTAL $F_1$ ♂♂ AND ♀♀ | $lz^+$ ♂♂ AND ♀♀ | $lz^g$-LIKE ♂♂ | PER CENT CROSSING-OVER* |
|---|---|---|---|---|
| $lz^{BS}/sn^3\ lz^{46}\ ras^4\ v$ | 20,554 | 9† | 5‡ | 0.09 |
| $lz^{BS}/ec\ ct^6\ lz^g\ v\ f$ | 16,255 | 13§ | 5‖ | 0.14 |
| $lz^{46}/ec\ ct^6\ lz^g\ v\ f$ | 16,098 | 4§ | 3‖ | 0.06 |

* See text for method of calculation.

† All carried $sn^3$. ‡ All carried $ras^4\ v$.

§ All carried $ct^6$. ‖ All carried $v$.

Three lozenge alleles were used in attempting to answer these questions: $lz^{BS}$, which occurred as a result of $X$-irradiation in $T(1{:}4)B^S$ and which has been separated from the translocation; $lz^{46}$, which is of spontaneous origin, and $lz^g$, which is the same mutant referred to above except that it has been extracted from *In(1)dl 49* where it originally occurred as a result of $X$-irradiation. All these mutants are typical lozenge alleles, recessive to wild type, and phenotypically characterized by derangement of the eye facets, reduction in the eye pigment and female infertility related to the absence of spermathecae.[2, 4, 5] All possible heterozygotes of the three alleles (e.g., $lz^{BS}/lz^{46}$) are lozenge in phenotype and difficult to separate phenotypically from either homozygote. Compounds among the three mutants were made and back-crossed to males of the genotype *In(1)dl 49, v* $lz^g$. The results are listed in table 1.

From the results tabulated in table 1, it may be noted that in the cross of females $lz^{BS}/sn^3\ lz^{46}\ ras^4\ v$ X males *In(1)dl 49, v* $lz^g$ (where $sn^3$ lies 6.7 units to the left and $ras^4$ and $v$ lie 5.1 and 5.3 units, respectively, to the right of lozenge) two unexpected classes of offspring occur. One class is wild type in appearance and carries the marker gene $sn^3$. These individuals are phenotypically identical to the $lz^+$ flies recovered from females $lz^g/lz^s$. The second class comprises individuals which carry the marker genes $ras^4$ and $v$ and are phenotypically completely separable from both $lz^{BS}$ and $lz^{46}$, and are in fact identical to $lz^s$ in phenotype. While $lz^+$ being dominant to both $lz^{BS}$ and $lz^{46}$ was recovered in both males and females, the $lz^s$-like individuals were males only. The failure to recover the $lz^s$-like chromosome in females stems from its phenotypic inseparability when compounded to $lz^{g3}$ consequently only one-half the expected number were found. This accounts for the discrepancy between the $lz^+$ and $lz^s$-like individuals in table 1 which theoretically should be equal in frequency.

It may be reasoned that the $lz^+$ flies carry an $X$-chromosome which as the result of crossing over carries wild-type alleles for both $lz^{BS}$ and $lz^{46}$, while the $lz^s$-like flies represent the complementary crossover with $lz^{BS}$ and $lz^{46}$ on the same $X$-chromosome. If this is the case, then it should be possible to recover both $lz^{BS}$ and $lz^{46}$ separately in the progeny of females of the genotype $lz^+/lz^s$-like. That this is the case may be determined from the following results. Females were constituted having one $X$-chromosome carrying $lz^s$-like (presumably $lz^{BS}$ and $lz^{46}$ together) plus the marker genes $ras^4$ and $v$, and the other an $lz^+$ $X$-chromosome (derived from crossing over between $lz^{BS}$ and $lz^{46}$) bearing the marker gene $sn^3$. Among 12,900 male offspring from these females, three males of the genotype $sn^3\ lz^{46}\ ras^4\ v$ and four males $lz^{BS}$ were recovered. These results support the hypothesis that the $lz^+$ chromosome carries wild-type lozenge alleles and the $lz^s$-like chromosome carries both $lz^{BS}$ and $lz^{46}$ and therefore might be designated as $lz^{BS,\ 46}$. That the $lz^+$ chromosome does contain only wild-type lozenge alleles is supported by one additional experimental finding. Females were constituted having one $X$-chromosome carrying $lz^{BS,\ 46}\ ras^4\ v$ and one chromosome derived from the Canton-$S$ wild-type stock and carrying the marker genes $sn^3$ and $v$. Among 9100 male offspring of these females, one male $sn^3\ lz^{46}\ ras^4\ v$ and three males $lz^{BS}\ v$ were recovered, thereby confirming the identity of the $lz^+$ chromosome with a wild-type $X$-chromosome and therefore as bearing only wild-type lozenge alleles.

The results reported thus far indicate that crossing-over is not peculiar to the alleles $lz^g$ and $lz^s$ since it occurs also between other alleles, $lz^{BS}$ and $lz^{46}$. They indicate also that the crossing-over between alleles is equal and not unequal as in the case of Bar, for the following reasons. First, in no instance among many thousands of flies examined have unexpected individuals ($lz^+$ or $lz^s$-like) been observed from females homozygous for $lz^{BS}$ or $lz^{46}$. Second, the crossover individuals resulting from females $lz^{BS}/lz^{46}$ are directed in so far as the marker genes are concerned. All $lz^+$ individuals carry the marker $sn^3$ while all $lz^{BS,\ 46}$ individuals carry $ras^4$ and $v$. If unequal crossing-over were involved then one would expect to get $lz^+$ flies carrying either $sn^3$ or $ras^4\ v$, and not always and only $sn^3$ as noted here. Similarly, $lz^{BS,\ 46}$ flies would be expected to carry either $sn^3$ or $ras^4$ and *v*, but not solely $ras^4$ and *v*. Consequently, it seems reasonable to conclude that regular crossing-over has occurred between two closely linked genes, with $lz^{BS}$ located just to the left of $lz^{46}$.

It may be noted in the cross of females $lz^{BS}/ec\ ct^6\ lz^g\ v\ f$ X *In(1)dl 49, v* $lz^g$ males (where *ec* and $ct^6$ lie 22.2 and 7.7 units, respectively, to the left and *v* and *f* lie 5.3 and 29 units, respectively, to the right of lozenge) that here too unexpected progeny of two types were obtained (cf. table 1). The $lz^+$ group is phenotypically identical to the $lz^+$ flies recovered from females $lz^{BS}/lz^{46}$, while the complementary crossover individuals, presumably carrying both $lz^{BS}$ and $lz^g$ on the same chromosome, are phenotypically inseparable from flies $lz^{BS,\ 46}$. As in the case of $lz^+$ flies from females $lz^{BS}/lz^{46}$, the $lz^+$ flies derived from females $lz^{BS}/lz^g$ all carry the left marker gene proximal to lozenge, namely $ct^6$. Similarly, the flies presumably $lz^{BS,\ g}$ carry the marker proximal to lozenge on the right, namely *v*.

These results indicate that the mutant $lz^g$ lies to the right of $lz^{BS}$. That $lz^g$ also lies to the right of the $lz^{46}$ and represents a third locus in the lozenge group of mutants may be readily deduced from the results obtained when females $lz^{46}/ec\ ct^6\ lz^g\ v\ f$ are crossed to males *In(1)dl 49, v* $lz^g$. It can be seen that once again two unexpected classes of progeny are obtained (cf. table 1). The $lz^+$ individuals are phenotypically identical to the $lz^+$ flies referred to previously and in this cross all carry the marker gene $ct^6$. The second class of unexpected individuals may again be described as $lz^s$-like and presumably carry both $lz^{46}$ and $lz^g$ on the same chromosome. In addition all such flies recovered carry the marker gene *v*. On the basis of the directed distribution of the marker genes to the $lz^+$ and the $lz^s$-like individuals, it seems logical to represent $lz^g$ as lying to the right of $lz^{46}$.

*Discussion*.—These observations may be fitted into a scheme which specifies that the lozenge allelic series is genetically divisible into at least three closely linked loci in the order of $lz^{BS}$–$lz^{46}$–$lz^g$, proceeding from left to right along the *X*-chromosome. The genetic interval separating $lz^{BS}$ and $lz^{46}$ is of the order of 0.09% crossing-over, and the interval separating $lz^{46}$ and $lz^g$ is of the order of 0.06% crossing-over. (The intervals have been calculated after first doubling the crossovers giving two lozenge genes in the same *X*-chromosome since, as noted previously, only one-half of these could be recovered.) If this interpretation is correct, then a wild-type *X*-chromosome may be represented as carrying three wild-type genes each allelic to one of the three lozenge mutant genes. With this as a basis, the *X*-chromosome bearing each of the mutant genes may be represented genotypically as follows: $lz^{BS}$ + +; + $lz^{46}$ +; and + + $lz^g$.

Phenotypic differences have been noted with combinations of any two of the lozenge mutants depending on whether the mutants are located together on the same $X$-chromosome or separately on homologous $X$-chromosomes. Using the genotypic notation proposed, the following combinations of mutant genes have been compared and their phenotypes recorded after each genotype:

$$lz^{BS} + +/+ \; lz^{46} + \text{ (mutant)}; \quad lz^{BS} \; lz^{46} +/+ + + \text{ (wild type)}$$
$$+ \; lz^{46} +/+ + \; lz^{g} \text{ (mutant)}; \quad + \; lz^{46} \; lz^{g}/+ + + \text{ (wild type)}$$
$$lz^{BS} + +/+ + \; lz^{g} \text{ (mutant)}; \quad lz^{BS} + lz^{g}/+ + + \text{ (wild type)}$$

It may be noted that in all cases the number of genes, mutant and wild type, is identical. Furthermore in all cases each mutant gene is balanced by a wild-type allele on the homologous $X$-chromosome. Yet when a lozenge mutant is present on each homologous $X$-chromosome, the wild-type alleles behave as recessive genes (or the lozenge mutants act as dominant genes), but when the same lozenge mutants are located together on the $X$-chromosome (even when separated by a wild-type allele) and the homologous $X$-chromosome carries only wild-type alleles, the wild-type alleles together act as dominant genes (or the lozenge mutants behave as recessives). These dominance relationships may be interpreted as being the result of a position effect. Similar observations have been made in the case of the interactions of the pseudoalleles Star and asteroid in *D. melanogaster*.[6]

A further conclusion which may be drawn in the light of additional information (unpublished) is that the three lozenge loci represent a reduplication of essentially identical genetic material. A number of other lozenge alleles (10 of the 14 which we possess) have been analyzed to determine to which locus they may be assigned.[7] While this analysis is incomplete, the data obtained thus far permit three generalizations which support this conclusion. In the first place, it is clear that the mutants at each locus possess the same array of phenotypic effects, vis., alteration of eye color and structure, infertility of females, etc. Secondly, if the mutants are classified quantitatively (e.g., with respect to the amount of red eye pigment formed), it can be seen that they are distributed at random to the three loci. Lastly, there appears to be no correlation between the mode of origin of the various mutants (i.e., spontaneous, x-ray induced) and their position. These findings are somewhat at variance with the observations made in cotton where three closely linked loci determining anthocyanin pigmentation have been described.[8, 9] In the latter case, although the three loci are in some instances interdependent, it cannot be said that all three possess the same array of phenotypic effects, suggesting that these loci represent three closely linked though not necessarily identical genes.

In addition to the cases cited, additional occurrences of crossing-over between what are apparently alleles have been observed in maize[10] and Drosophila,[11] and possibly in the case of the Brachy series in mice.[12] Just how these cases relate to the lozenge series is not at the present time completely clear, but it may be that all represent further examples of duplication or reduplication of the same genic material. It is not possible at present to extrapolate the data of the lozenge series to other allelic series. However, they do not justify the conclusion that all multiple allelic series represent cases of multiple loci, for the simplest explanation for the situation within the lozenge complex itself would appear to be that of three closely linked multiple allelic series.

*Summary.*—1. Crossing-over has been observed between three sex-linked, recessive lozenge alleles, $lz^{BS}$, $lz^{46}$, $lz^{g}$ in *D. melanogaster.*

2. From females heterozygous for any two of the mutants, wild type *X*-chromosomes and *X*-chromosomes bearing two lozenge mutants have been recovered.

3. On the basis of the crossover results, the lozenge allelic series may be subdivided into three closely linked loci.

*Acknowledgment.*—The authors are indebted to Drs. C. P. Oliver, H. J. Muller and E. Novitski for making available mutants used in these experiments.

[1] Oliver, C. P., Proc. Natl. Acad. Sci., **26**, 452–454 (1940).

[2] Oliver, C. P., and Green, M. M., *Genetics*, **29**, 331–347 (1944).

[3] Sturtevant, A. H., *Ibid.*, **10**, 117–147 (1925).

[4] Oliver, C. P., *U. Texas Publ.*, **4720**, 167–184 (1947).

[5] Green, M. M., *Am. Natl.*, **82**, 188–192 (1948).

[6] Lewis, E. B., *Genetics*, **30**, 137–166 (1945).

[7] Green, M. M., and Green, K. C., unpublished.

[8] Stephens, S. G., *Genetics*, **33**, 119–214 (1948).

[9] Yu, Chi Pao, and Chang, Te Sung, *J. Genetics*, **49**, 46–56 (1948).

[10] Laughnan, J. R., Proc. Natl. Acad. Sci., **35**, 167–178 (1949).

[11] Lewis, E. B., *Genetics*, **33**, 113 (1948).

[12] Dunn, L. C., and Caspari, E., *Ibid.*, **30**, 543–568 (1945).

---

# EXPERIMENTS ON LIGHT-REACTIVATION OF ULTRA-VIOLET INACTIVATED BACTERIA

By A. Novick and Leo Szilard

Institute of Radiobiology and Biophysics, University of Chicago

Communicated by H. J. Muller, August 15, 1949

Many types of microorganisms are killed by ultra-violet light and the number of survivors falls off with increasing dose. A. Kelner[1] reported recently on his discovery that, if exposure to ultra-violet light is followed by exposure to visible light, the number of survivors is very much larger. A similar discovery was reported by R. Dulbecco[2] for bacteriophage.

We investigated this phenomenon of light-reactivation of ultra-violet inactivated bacteria and found some very simple regularities.

In our experiments we used a strain of coli, *B/r*, originally isolated by E. M. Witkin.[8] Cultures were grown in a lactate-ammonium phosphate medium under aeration to between $5 \times 10^7$ and $10^8$ bacteria per cc., then transferred into saline, and incubated under aeration at 37°C. from 14 to 18 hours. The cultures were then placed in the icebox, kept at about 6°C. and used for experiments over a period of about one week.

Figure 1 shows in semilogarithmic plot the result of one series of experiments. Eight other such series were conducted with similar results.

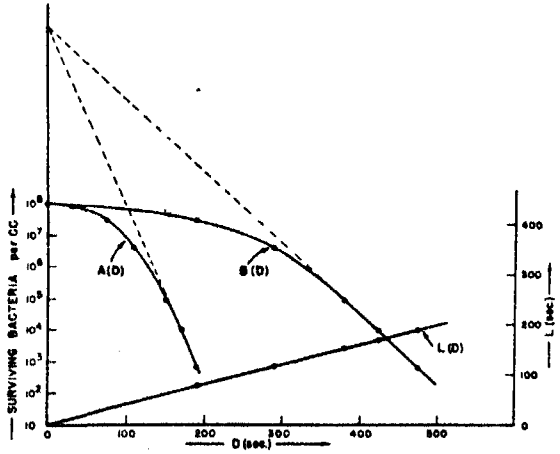


FIGURE 1

Experiment of May 11, 1949. $A(D)$, left scale, gives in semi-log plot the number of survivors as function of ultra-violet dose, $D$, in seconds of exposure to 15-watt germicidal lamp at 50 cm. distance (no light reactivation). $B(D)$, left scale, gives in semi-log plot the number of survivors as function of ultra-violet dose, $D$, with light reactivation by 1-hour exposure to light of 1000-watt projection lamp at 8 inches distance. $L(D)$, right scale, gives in linear coordinates $L$ as function of $D$ satisfying equation $B(D) = A(L)$.

In these series of experiments the survivor curves, $A(D)$, were obtained in the following manner: The bacteria were exposed in saline to different ultra-violet doses, $D$, characterized by the number of seconds of exposure, then plated on nutrient broth agar, incubated, and the survivors determined by colony count. A 15-watt germicidal lamp was used as the source of ultra-violet radiation with the mercury line at 2536 Å. predominating in the spectrum. A 15-cm. Petri dish containing 20 cc. of the bacterial suspension was exposed at a distance of 50 cm. from the lamp and was agitated during exposure.

The survivor curves, $B(D)$, were obtained in a similar manner, except that immediately following exposure to the ultraviolet lamp the bacterial suspensions were exposed for one or two hours, at 8-inch distance, to the light of a 1000-watt projection lamp, and then plated on nutrient broth agar. The color temperature of the lamp was close to 3350°K. The bacterial suspensions were kept at 37°C. during exposure to the light of the projection lamp. In most of the series a 1-inch layer of a 0.05 molar solution of $CuCl_2$ was interposed between the lamp and the bacterial suspension.

As can be seen in figure 1, in semilogarithmic plot, both survivor curves $A$ and $B$ drop with increasing ultra-violet dose, $D$, slowly at first, then faster, and finally go over into a straight line.

As figure 1 shows, light-reactivation is a striking phenomenon. For an ultra-violet dose of $D = 190$ sec., for instance, the number of survivors falls in the absence of light-reactivation (curve $A$) from $10^8$ to 700. With light reactivation, however, (curve $B$) the number of survivors only falls from $10^8$ to $3 \times 10^7$; i.e., due to light-reactivation the titre of viable bacteria increases by a factor of about 40,000. In order to obtain the same drop in the survivors (from $10^8$ to $3 \times 10^7$), in the absence of light-reactivation, a *lower* ultra-violet dose of $D = 75$ sec. would have to be employed. For every ultra-violet dose, $D$, which is followed by light-reactivation and thus leads to a certain number of survivors, $B(D)$, there can be found a lower ultra-violet dose, $L$, which in the absence of light-reactivation would lead to the same number of survivors; i.e., for which we have $B(D) = A(L)$.

Plotting $L$ as a function of $D$ in figure 1 we obtain for $L(D)$ a straight line going through the origin, for which we may write $L(D) = qD$, where $q$ is a dose-independent constant smaller than 1. This means that there is a simple relationship between the survivor curves $A$ and $B$ which we may express by writing

$$B(D) = A(qD). \qquad (1)$$

This relationship is confirmed by the nine series of experiments that were carried out. In any one of them the bacteria were light-reactivated by an exposure to the light of a projection lamp, at 8-inch distance, either of one hour or of two hours. After exposure of two hours the number of survivors obtained is, within the limits of experimental error, indistinguishable from the maximum reactivation that can be obtained with longer exposures to light. After exposure of one hour the number of survivors is lower, but the difference is not great. The $q$ values obtained in these experiments vary slightly from series to series within the range of $0.32 < q < 0.42$. In the experiments shown in figure 1 we have $q = 0.4$.

If the light intensity used for light-reactivation is doubled by placing the bacterial suspension closer to the projection lamp, the maximum number of survivors obtainable remains the same, only it is obtained with a shorter exposure to the light.

From the relationship (1) it follows that if we extrapolate in the semi-logarithmic plot the straight line portions of the survivor curves $A$ and $B$ to zero ultra-violet dose, we must be led back to the *same point.* This is shown by the dotted straight lines in figure 1.

In order to interpret our results we assume for the *sake of argument* that when bacteria are irradiated with ultra-violet light a "poisonous" chemical compound, $P$, is produced in an amount which is *proportionate* to the ultra-violet dose, $D$. We assume that this "poisonous" compound is produced in two forms: a form $P_x$, which is not sensitive to light and which is produced in the amount $x_0$, and a form $P_y$, which can be destroyed by light and which is produced in the amount $y_0$, so that $D = x_0 + y_0$. We further assume that the ratio $x_0/y_0$ is independent of the ultra-violet dose $D$. And finally we assume that the number of survivors after exposure to ultra-violet irradiation, with or without subsequent light-reactivation, is determined by the amount of "poison" that is present in the bacteria *at the time* they are incubated with nutrient medium and permitted to multiply.

On the basis of these assumptions, we may now account for the relationship (1) between the survivor curves $A$ and $B$ by saying that if we follow up the ultra-violet irradiation of a bacterial suspension by exposure to strong light for one or two hours we destroy all the poison present in the form $P_y$, and leave only the poison $P_x$, and by further saying that the amount of the poison present in the form $P_x$ is given by $x_0 = qD$, where $q$ is a dose-independent constant. The amount of "poison" left in the bacteria after light-reactivation is then given by $qD$, and therefore the number of survivors after light-reactivation, $B(D)$, can be taken from the survivor curve obtained in the absence of light-reactivation, $A(D)$, by writing $B(D) = A(qD)$.

If we now further assume that the light-sensitive variety of the poison, $P_y$, which is present in an amount $y$, is destroyed by light at a rate which is proportionate to its amount, $y$, we may write $\frac{dy}{dt} = -\alpha y$, where $\alpha$ is a function of the light intensity and independent of the dose, $D$. From this assumption it follows that if a bacterial suspension is exposed to an ultra-violet dose, $D$, and immediately afterward is exposed to strong light (say, to the light of a 1000-watt projection lamp at 8-inch distance) for a period of time, $t$, the total amount of poison, $L = x + y$, present in the bacteria is at any time, $t$, given by

$$L = x_0 + (D - x_0)e^{-\alpha t}. \qquad (2)$$

For an infinitely long exposure to light, we have $L_\infty = x_0$. We may then write

$$\ln(L - L_\infty) = \ln(D - L_\infty) - \alpha t. \qquad (3)$$

The result of an experiment undertaken to check this relationship (3) is shown by curve I in figure 2. In this experiment we first exposed the bacterial suspension to the ultra-violet lamp for $D = 250$ sec. and then to the light of a 1000-watt projection lamp at 8-inch distance. We determined the number of survivors as a function of $t$, the time of exposure to the projection lamp, by plating on broth agar from aliquots taken from the bacterial suspension at different times, $t$, up to a time of 2 hours. From the number of survivors, we computed by means of the survivor curve, $A(L)$, the corresponding $L$ values. As $L_\infty$ we took the $L$ value corresponding to the number of survivors obtained after 2 hours of exposure to the projection lamp.

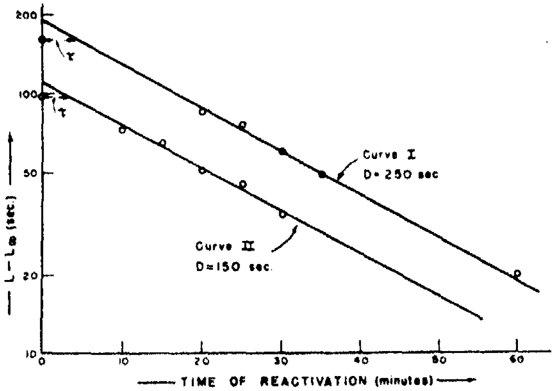


FIGURE 2

Experiment of June 23, 1949. Curve I gives in semi-log plot $L - L_\infty$ as a function of time of reactivation at 8 inches from 1000-watt projection lamp for bacteria that had received an ultra-violet dose, $D$, of 250 sec. Curve II is the same for bacteria that had received an ultra-violet dose of 150 sec. The interval $\tau$ represents the latent period of the process of reactivation.

As can be seen from curve I in figure 2 the points obtained fall on a straight line in accordance with equation (3). This type of relationship was confirmed in five series of experiments, giving values of $\alpha$ ranging from 2.19 to 2.38 per hour.

It should be noted however that the straight line I in figure 2 does not extrapolate back to the experimental point for zero exposure to light and the discrepancy corresponds to a latent period, $\tau$, of about 3 minutes.

Line II in figure 2 differs from line I only in so far as it was obtained with an exposure to the ultra-violet lamp of $D = 150$ seconds. The fact that these two lines are parallel is in accord with the thesis that $\alpha$ is independent of the ultra-violet dose, $D$. The latent period $\tau$ manifested by line II is again about 3 minutes. Two series of experiments were performed to demonstrate in this manner the independence of $\alpha$ of the ultra-violet dose $D$.

Figure 3 shows the result of three experiments in which light-reactivation was carried out by exposing the ultra-violet irradiated bacterial suspension to the light of the projection lamp for $t = 20$, 25 and 30 min., respectively. From the number of survivors observed we computed, by means of the survivor curve, $A(D)$, the corresponding $L$ values, and in each case we obtained for $L(D)$, as can be seen from figure 3, a straight line passing through the origin.

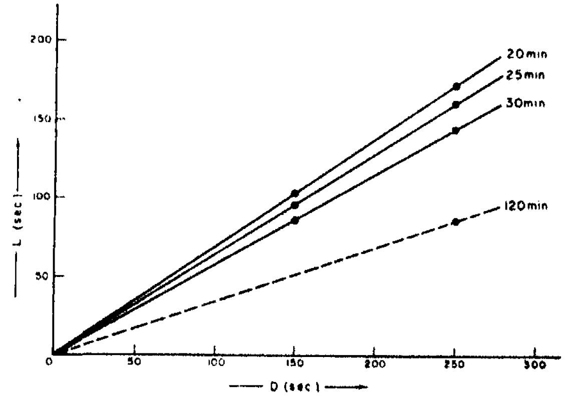


FIGURE 3

Experiment of June 23, 1949, gives $L$ as a function of the ultra-violet dose, $D$, for 20, 25 and 30 minutes of light reactivation at 8 inches from 1000-watt projection lamp.

This result is to be expected provided that both $\alpha$ and $\tau$ are independent of the ultra-violet dose, $D$. Two such series of experiments were performed with identical results.

We were also interested to see in what manner the value of $\alpha$ depends on the intensity of the light which is used for reactivation. We varied this intensity by placing the bacterial suspension at different distances from the projection lamp. In such experiments we found that the value of $\alpha$ decreased with decreasing light intensity, but the decrease was less than strict proportionality with intensity would lead one to expect. The latent period $\tau$ increased in these experiments with decreasing light intensity.

We did not, however, in any of these experiments lower the light intensity very much below the value that obtains at 8inches from the projection lamp. We kept the intensity comparatively high in order to be able to neglect a reaction which takes place in the dark and therefore, of necessity, must also manifest itself if light-reactivation is carried out by means of very weak light. This dark reaction is easily observed if the bacterial suspension is irradiated with ultra-violet and then incubated in the dark at 37°C., for instance for 3 hours, prior to being reactivated at 8-inch distance from the projection lamp. The number of survivors after light-reactivation is then found to be much lower than is obtained when no dark incubation is interposed between ultra-violet inactivation and light-reactivation.

This concludes the report on our findings concerning the light-reactivation of ultra-violet inactivated bacteria in so far as these findings relate to the effect of light-reactivation on the number of survivors.

It is known that ultra-violet radiation not only kills bacteria but also produces mutants among the progeny of the survivors. We were interested to find out how light-reactivation affects the number of mutants which appear among these progeny. Particularly convenient for this purpose are mutants which manifest resistance to a bacteriophage, and we chose for the purpose of our investigation mutants resistant to the coli phages $T4$, $T6$ or $T1$.

M. Demerec[4] had studied the ultra-violet-induced mutations in coli to resistance to phage $T1$ and discovered that the great majority of the mutants do not appear immediately following ultra-violet irradiation, but that most of the bacteria have to go through several cell divisions before their mutation becomes phenotypically expressed.

A small minority of the mutants is phenotypically expressed prior to any cell division and A. Kelner finds that the appearance of such mutants is suppressed if the ultraviolet irradiation is followed by light-reactivation (see A. Kelner, abstracts of papers presented at the 49th General Meeting of the Society of American Bacteriologists, page 14, May 1948).

In our experiments we allowed, following irradiation, the survivors to go *through ten generations in liquid culture* and determined among the progeny the number of phage-resistant mutants per $10^8$ bacteria.

If we are right in assuming that a poison is produced by the ultra-violet rays employed in our experiments, that the amount of this poison is reduced by light-reactivation, and that the amount of poison which is present in the bacteria when they are allowed to multiply determines the number of survivors, it then seems possible that this same "poison" might also determine the number of mutants—resistant to one of the phages—that appears among the progeny of the survivors. In this case we should expect that the number of mutants appearing among the progeny of the bacteria which were given a certain ultra-violet dose, $D$, and which were subsequently light-reactivated will be the same as the number of mutants appearing among the progeny of the bacteria which were exposed to a *lower* ultra-violet dose, $qD$, but which were *not* light-reactivated.

In our experiments the bacteria were exposed in saline suspension to different ultra-violet doses, $D$. Aliquots were then taken which were either incubated with lactate-ammonium phosphate medium at 37°C. under aeration or they were first exposed to the light of the projection lamp at 8-inch distance for one hour and were subsequently thus incubated. In either case the bacteria were allowed to go through about *ten generations* in liquid culture and were then assayed for mutants resistant to one of the coli phages $T4$, $T6$ or $T1$.

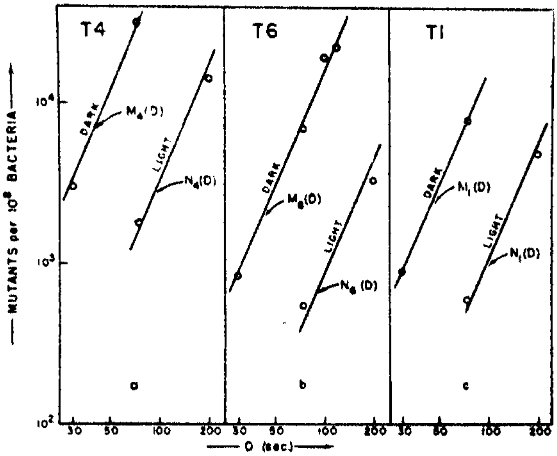


FIGURE 4 (A), (B) AND (C)

Experiment of July 2, 1949, gives in log-log plot the number of phage-resistant mutants per $10^8$ bacteria as a function of the ultra-violet dose $D$ both without (*DARK*) and with (*LIGHT*) light reactivation for phages $T4$, $T6$ and $T1$ in figures $a$, $b$, and $c$, respectively. The bacteria went through 10 generations in liquid culture prior to assaying for the mutants. The values shown in the graph are corrected for the spontaneous mutants which amounted to 800, 40 and 170 per $10^8$ bacteria for $T4$, $T6$ and $T1$, respectively.

Figure 4 shows the result of one series of experiments in log log plot. The symbols $M_4(D)$, $M_6(D)$ and $M_1(D)$ relate to bacteria which were exposed to ultra-violet, were *not* light-reactivated, and went through ten generations in liquid culture (corresponding to an increase in bacterial titre by a factor of about 1000). Similarly $N_4(D)$, $N_6(D)$ and $N_1(D)$ relate to bacteria which were exposed to ultra-violet, were light-reactivated, and went through ten generations in liquid culture. These symbols designate the number of mutants resistant to phages $T4$, $T6$ and $T1$, respectively, which were found per $10^8$ bacteria in the cultures—corrected for the spontaneous mutants that were obtained for zero ultra-violet dose—and plotted as a function of the ultra-violet dose, $D$.

Within the range of this experiment and within the limits of its accuracy, the lines $M(D)$ and $N(D)$ are represented in the log log plot by straight lines which are *all parallel* to each other. For each phage and for any given ultra-violet dose, $D$, the value of $M$ is higher than the value of $N$. For every ultra-violet dose, $D$, which leads, when followed by light-reactivation, to a certain number of mutants, $N(D)$, there can be found a *lower* ultra-violet dose, $L$, which in the absence of light-reactivation leads to the same number of mutants; i.e., for which we have $N(D) = M(L)$.

Because the lines $M(D)$ and $N(D)$ are represented by *parallel* straight lines in the log log plot we may write:

$$N_4(D) = M_4(m_4D); \qquad N_6(D) = M_6(m_6D); \qquad N_1(D) = M_1(m_1D) \qquad (4)$$

where $m_4$, $m_6$ and $m_1$ are dose-independent constants. The values of $m_4$, $m_6$ and $m_1$ taken from the experimental series plotted in figure 4 are 0.3, 0.3 and 0.32, respectively. For the same series of experiments the value of $q$ was also determined and was found to be 0.35. Within the limits of our experimental accuracy we have thus found: $m_4 = m_6 = m_1 = q$. Two other series of experiments which were undertaken also gave, within the limits of experimental error, identical values for $m$ and $q$.

It should be noted, however, that our experiments establish the validity of relationship (4) with much less accuracy than they establish the validity of relationship (1). In general, in our experiments the determination of the number of the mutants in any one single measurement was affected by an error of ±50%, whereas the determination of the number of survivors in any one single measurement was affected only by an error of ±20%. And more important, due to the shape of the survivor curve $A(D)$, a large percentage error in the number of survivors leads only to a small percentage error in the corresponding $L$ values taken from the survivor curve $A$, whereas an error of 50% in the number of mutants, according to the curves $M$ and $N$ plotted in figure 4, leads to an error in the corresponding ultra-violet dose, $D$, of about 25%.

We have not investigated the ultra-violet-induced mutations to phage resistance with the same thoroughness as we have the effect of ultra-violet exposure on the number of survivors. Therefore at this time we can only say that our results to date are entirely consistent with the view that the effect of light-reactivation on the appearance of mutants among the progeny of the ultra-violet-irradiated bacteria is the same as is its effect on the number of survivors and that this effect consists in the reduction of the effectiveness of the ultra-violet dose by the dose-independent factor $q$. This makes it possible to surmise that in our experiments the killing of the bacteria and the production of the mutants might be due to the *same* chemical effect produced by the ultra-violet irradiation.

[1] Kelner, A., PROC. NATL. ACAD. SCI., **35**, 73 (1949).

[2] Dulbecco, R., *Nature*, **163**, 949 (1949).

[3] Witkin, E. M., *Genetics*, **32**, 221 (1947).

[4] Demerec, M., PROC. NATL. ACAD. SCI., **32**, 36 (1946); Demerec, M., and Latarjet R., *Cold Spring Harbor Symposia*, **11**, 38 (1946).

---

# SOME TWO-PERSON GAMES INVOLVING BLUFFING*

BY RICHARD BELLMAN AND DAVID BLACKWELL

STANFORD UNIVERSITY AND HOWARD UNIVERSITY

Communicated by J. von Neumann, August 4, 1949

1. *Introduction*.—It is now possible, with the aid of the von Neumann theory of games, to give a systematic theoretical treatment of the two-person zero-sum game. The techniques of this modern theory have been shown to have wide and important applications in such fields as economic theory, the theory of statistics and so on. Unfortunately, although the foundations of the theory are now classic, the actual solution of any particular game is generally a matter of some difficulty, involving a combination of frontal attack, applying routine methods, and a type of mathematical ingenuity that has been described as "low cunning."

Consequently, it is of interest to present a class of games, which may be handled by uniform techniques, in the hope that an accumulation of such examples will eventually render visible the mechanism behind the particular solutions, hitherto discovered only by artifice. Such has been the history of the theory of differential equations, the theory of probability and many of the other components of the mathematical edifice.

Consider the following two-person game where we call one player $B$, for bettor, and the other $D$, for dealer. Before play begins both players ante one. $D$ then deals a card $x_1$ to $B$, where $x_1$ is a random number in the interval $[0, 1]$ having the distribution function $F_1$, and a card $y_1$ to himself, $y_1$ also belonging to the interval $[0, 1]$ with distribution function $G_1$. $B$ initiates the betting, with the alternate possibilities of folding, in which case $D$ wins the ante, or of betting an amount $f(x)$, $1 \leq f(x) \leq M$. Once $B$ has bet, $D$ has a choice of folding, in which case $B$ wins the ante, of covering $B$'s bet or of raising an amount $g(f(x), y)$, $1 \leq g \leq M$. If $D$ raises, $B$ has a choice of covering or raising and so on, for at most $N_1$ raises by either player.

After the initial betting, $D$ deals two more cards, $x_2$ to $B$ and $y_2$ to himself, and again $B$ initiates the betting. This continues for at most $N_2$ draws. At the end of the betting, the hands are compared and the stronger hand wins the total amount wagered. The strengths of the hands are certain functions, $S(x_1, x_2, \ldots, x_n)$, $T(y_1, y_2, \ldots, y_n)$ of the cards, $x_i$, $y_i$. For the case of a deal and one draw, the above describes a poker game with two players; the case of several draws is analogous to Stud Poker; the case of several draws, but not simultaneously by both players, with one stage of betting and no raises, corresponds to Blackjack, sometimes called Twenty-One, with its arithmetically simpler cousin, Seven-and-a-Half; the case of no draws with the dealer always covering the bet is illustrated by Red Dog, and clearly many other common card games are described in many essentials by the above game.

Consequently, it is of some interest to investigate the problem of determining the best method of play for both players, where by "best" we mean the method that yields a maximum expectation.

The general problem seems quite difficult. In this paper we treat the simplified game where no draws are made after the initial deal, where there are only a finite number of bets allowed, $1 \leq z_1 < z_2 < \ldots < z_n$, with no raises, and where the distribution functions, $F$ and $G$, are both equal to $x$.

However, it seems probable that the methods we present to treat the simplified version will be applicable to the $N$-draw game, and we hope to investigate the general question in the future.

That our procedures are applicable to the continuous case, and not to the discrete case, depends upon the result which is stated precisely below, that in certain continuous games only pure strategies are necessary, whereas the discrete game, in general, requires mixed strategies. Here the terms "pure" and "mixed" are used in the sense of game theory, and have no connection with the possible motives of either player. Given a card $x$, there are always several alternate procedures or strategies for either player to follow. A pure strategy is one which compels him to follow a unique course of action whenever he receives a card $x$; a mixed strategy is one which furnishes a rule for mixing the alternate available courses of action in certain proportions determined by $x$. A pure strategy then is a limiting form of a mixed strategy.

This result, that only pure strategies are required in the continuous case, seems quite unexpected, and at first sight to run counter to the known (heuristic) theory of bluffing. However, the paradox is easily resolved. In a continuous game, one allows the game, which furnishes a random card, to do the bluffing. It turns out that if the cards are dealt at random, any further randomization furnished by mixed strategies on the part of the players is superfluous.

This is not to say that there is no bluffing. Rather, one bluffs uniformly. This will be seen by examining the solution of the one-deal, no-draw game presented in the following paragraphs.

To the best of our knowledge, there have been only two previous mathematical discussions of the game of poker, the first by Borel,[1] and the second by von Neumann.[2] Borel's treatment depends more upon probabilistic considerations and involves no systematic use of game theory. One of the games discussed by von Neumann involves a choice of two bets for *B*, high or low (but no ante), and a choice for *D*, after having seen *B*'s bet, of seeing or not. The game described in paragraph 3, following, differs from this only in allowing *B* the additional privilege of folding, at a cost, however. The solution obtained by von Neumann is, as in our case, a pure strategy for both players, with bluffing for *B*. Our methods, which lean heavily upon Theorem 1 below, are different from those of von Neumann.

2. *A Fundamental Theorem.*—The solution of a game is often simplified if it is known in advance that one of the players has a good pure strategy. Even if it is known in advance only that every mixed strategy for player *I* can be approximated by pure strategies, the value of the game can still be found by solving the restricted game in which both players are limited to pure strategies, but *II* discovers *I*'s strategy. If in the restricted game *I* has a good strategy, this is also a good strategy in the original game.

There is a simple condition which sometimes ensures that mixed strategies may be approximated by pure strategies: if, at the beginning of the game, there is a chance move whose outcome is known to only one of the players, it may happen that he can use this chance move as a substitute for mixing strategies, so that a formally pure strategy, depending in an intricate way on the outcome of the chance move, can be constructed to approximate any given mixed strategy. This principle guarantees the existence of such approximating pure strategies in the one deal-no draw case of the game described above, and most probably in the general case too, although we have not as yet investigated this. That the actual good strategies are pure is not guaranteed by our principle, and, so far as we see, could not have been predicted in advance.

For our purpose, the relevant features of the game are the following, from *I*'s point of view.

1. At the beginning of the game, there is a chance move consisting of the selection of a number $x$ at random from the unit interval. $x$ is known to *I* but not to *II*.

2. *I*'s strategy is the choice of a number $i = 1, \ldots, N$, depending on $x$, i.e., a function $i(x)$, assuming only values $1, \ldots, N$.

3. The outcome $w(x, i; z)$, where $z$, *II*'s strategy, may now vary over an arbitrary space, is for fixed $i$ and $z$, a function of $x$ consisting of $R$ (fixed) continuous pieces, the modulus of continuity of the pieces being

uniform over $(i, z)$. (In the poker game described, $w$ consists of two constant pieces over the sets $x < y$, $x > y$.) Moreover, $w$ is uniformly bounded; say $|w| \leq M$.

The payoff to player $I$, i.e., the expected outcome for fixed pure strategies of $I$ and $II$, is then

$$v(i(x);\ z) = \int_0^1 w(x, i(x);\ z)dx. \tag{1}$$

THEOREM 1. *In a game satisfying 1, 2, 3, for every mixed strategy $F$ for $I$ and every $\epsilon > 0$, there is a pure strategy $i(x)$ such that, for every $z$,*

$$v(F;\ z) \leq v(i(x);\ z) + \epsilon. \tag{2}$$

*Proof.* A mixed strategy $F$ for player $I$ is a choice of $N$ functions $p_1(x), \ldots, p_n(x)$, $p_i(x) \geq 0$, $\sum_{i=1}^{N} p_i(x) = 1$ for each $x$, $p_i(x)$ specifying the probability he chooses $i$ with a given $x$. Then $v(F;\ z) = \sum_{i=1}^{N} \int_0^1 w(x, i;\ z) p_i(x)dx$. Divide the interval $0 \leq X \leq 1$ into disjunct subintervals $I_1, \ldots, I_n$, $m(I_j) < \delta$, where $\delta$ is chosen so that each continuous piece of each $w(x, i, z)$ varies by less than $\epsilon$ over any interval of length $\delta$. Now divide $I_j$ into $N$ intervals $I_{j1}, \ldots, I_{jN}$ so that

$$m(I_{jk}) = \int_{I_j} p_k(x)dx. \tag{3}$$

Then the pure strategy $i(x) = k$ on $I_{jk}$, $j = 1, \ldots, n$, $k = 1, \ldots, N$ is the required approximation to $F$. For

$$D = v(F;\ z) - v(i(x);\ z) = \sum_{j=1}^{n} \sum_{k=1}^{N} D_{jk}, \tag{4}$$

where

$$D_{jk} = \int_{I_j} w(x, k;\ z)p_k(x)dx - \int_{I_{jk}} w(x, k;\ z)dx.$$

Hence, $|D| \leq \sum_{j \epsilon C} \sum_{k=1}^{N} |D_{jk}| + 2MNR\delta$, where $C$ contains those $j$'s for which every $w(x, k, z)$ for the given $z$ is a single piece over $I_j$. The bound $2MNR\delta$ follows from the fact that there are at most $NR$ $j$'s not in $C$. For $j \epsilon C$,

$$|D_{jk}| \leq \left| U_{jk} \int_{I_j} p_k(x)dx - L_{jk} \int_{I_{jk}} dx \right| \leq \epsilon\, m(I_{jk}), \tag{5}$$

where $U$, $L$ are the upper and lower bounds of $w(x, k, z)$ over the interval $I_j$. Then

$$|D| \leq \sum_{j \epsilon C} \epsilon\, m(I_j) + 2MNR\delta \leq \epsilon + 2MNR\delta. \tag{6}$$

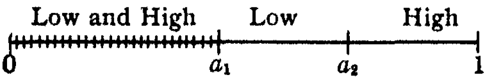
Since $\delta$ is arbitrarily small, the proof is complete.

3. *The Solution of a One-Draw Poker Game.*—In this section we present

the solution to the one-draw poker game with no raises and two bets of magnitude $z_1$, $z_2$, $z_2 \geq z_1 \geq z_0$, a constant specified below, the size of the ante being one. Complete details and discussion of the general case involving raises and a wider selection of wagers will be presented subsequently.

The solution shows that $D$'s strategy is unique and involves no bluffing, in the sense that he never calls with a weak hand, while $B$'s strategy is semi-determinate and involves one type of bluffing, betting high in low hands but not the other type, betting low on high hands.

$B$'s strategy is shown diagrammatically by the following decomposition of the $x$ interval [0, 1].



(1)

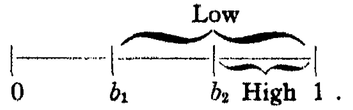
where

$$\begin{aligned} a_1 &= 1 - c_1 \\ a_2 &= 1 - c_1c_2/2, \end{aligned} \tag{2}$$

(where we have set $c_1 = 2/2 + z_1$, $c_2 = 2/2 + z_2$).

$B$ bets high if $a_2 \leq x \leq 1$, low if $a_1 \leq x \leq a_2$ and if $0 \leq x \leq a_1$ must bet high $H$ on a set of measure $(1 - c_2)c_1c_2/2$, low on a set $L$ of measure $(1 - c_1)$ $(c_1 - c_1c_2/2)$, and fold on the remaining part of $[0, a_1]$. The sets $H$ and $L$ are of fixed measure, but their location is arbitrary within $[0, a_1]$.

$D$'s strategy is exhibited by



(3)

where $b_1 = a_1$, $b_2 = 1 - c_2$. $D$ folds if $0 \leq y \leq b_1$, sees a low bet if $y \geq b_1$, and a high bet if $y \geq b_2$. From this it is clear why the exact location of the sets on which $B$ bets high or low when he has a card in the interval $[0, a_1]$ is of no importance. The value of the game is

$$v = -(1 - c_1)^2 + c_1c_2(c_1 - c_2), \tag{4}$$

which is always negative for $0 \leq c_2 \leq c_1 \leq c_0$, where $c_0$ is the smallest root of

$$c^3 - 8c^2 + 16c - 8 = 0, \tag{5}$$

$c_0 \simeq 0.76$. The constant $z_0$ mentioned above is $2\left(\frac{1 - c_0}{c_0}\right) \simeq 0.62$. Thus, we see the character of the solution changes at the point where the game becomes fair to $B$. We have not as yet investigated the game in the case $z_1 < z_0$, since the most interesting case is $z_2 > z_1 \geq 1$ (the amount of the ante).

* The results contained in this paper were obtained in the course of research conducted for The Rand Corporation, Santa Monica, California.

[1] Borel, E., and collaborators, *Traite du Calcul des Probabilités et de ses Applications*, Vol. IV, 2, Chapter V, Paris, 1938.

[2] von Neumann, J., and Morgenstern, O., *The Theory of Games and Economic Behavior*, Chapter 19, Princeton University Press, Princeton, 1947.

---

# ON FLUCTUATIONS IN COIN-TOSSING*

BY KAI LAI CHUNG AND W. FELLER

CORNELL UNIVERSITY, ITHACA, NEW YORK

Communicated by J. von Neumann, August 4, 1949

In the classical coin-tossing game we have a sequence of independent random variables $X_\nu$, $\nu = 1, 2, \ldots$ each taking the values $\pm 1$ with probability 1/2. We are interested in the signs of the partial sums $S_n = \sum_{\nu=1}^{n} X_\nu$. To eliminate the zeros of $S_n$ we make the following convention: $S_n$ is "positive" if $S_n > 0$ or if $S_n = 0$ but $S_{n-1} > 0$; otherwise $S_n$ is "negative." The elegance of the results to be announced depends on this convention. Let $N_n$ denote the number of "positive" terms among $S_1, \ldots, S_n$. We shall confine ourselves to an even $n$, noting only that $0 \leq N_{n+1} - N_n \leq 1$. In the following $r$ and $m$ are positive integers and $P(B|A)$ is the conditional probability of $B$ under the hypothesis $A$.

THEOREM 1.

$$P(N_{2n} = 2r) = \frac{1}{2^{2n}} \binom{2r}{r} \binom{2n-2r}{n-r}.$$

THEOREM 2A.

$$P(N_{2n} = 2r | S_{2n} = 0) = \frac{1}{n+1}.$$

THEOREM 2.

$$P(N_{2n} = 2r | S_{2n} = 2m) = m \binom{2n}{n-m}^{-1} \sum_{i=m}^{r} \binom{2i}{i-m} \binom{2n-2i}{n-i} \frac{1}{i(n-i+1)}.$$

From these theorems it is easy to derive the following *limiting forms*

$$\lim_{n \to \infty} P(N_n \leq \alpha n) = \frac{2}{\pi} \text{arc sin } \alpha^{1/2} \qquad 0 < \alpha < 1 \qquad (a)$$

(with the corresponding density $\pi^{-1}(\alpha - \alpha^2)^{-1/2}$), and

$$\lim_{n \to \infty} P(N_n \leq \alpha n | S_n = \lambda n^{1/2}) =$$
$$\pi^{1/2} \int_{(1-\alpha)\lambda^2/\alpha} (\lambda^2(\alpha - 1) + \alpha y) y^{-3/2} e^{-y}\, dy \qquad (b)$$

for $0 < \alpha < 1$ and for every fixed $\lambda$ such that $\lambda n^{1/2}$ is a possible value of $S_n$. As $\lambda \to 0$ the last integral tends to $\alpha$ and for $\lambda = 0$ we get in the limit the uniform distribution as suggested by Theorem 2A.

The limiting form ($a$) of Theorem 1 is contained in a result of Erdös and Kac.[1] An analog to Theorem 1 for continuous random variables was proved by E. S. Andersen;[2] however, his result does not hold for arbitrary discrete variables, and Theorem 1 is not contained in Andersen's result.

The most surprising information is contained in Theorem 2A. To see this, note that the distribution of Theorem 1 departs radically from the familiar bell-shaped pattern. The limiting probability density curve resulting from Theorem 1 is U-shaped with the mean at the minimum. More precisely, while $\lim_{n \to \infty} E(N_{2n}/2n) = \frac{1}{2}$ this is the *least* probable value of $N_{2n}/2n$. This means that although the game is a symmetrical one (hence "fair") it is nevertheless more likely that one party "leads" in an overwhelmingly large proportion of the time than that each party leads about half of the time. In Theorem 2A it is plausible that the knowledge that $S_{2n}$ vanishes should reduce the probability of extreme values of $N_{2n}$. It is noteworthy that this is true to the extent of a *uniform* distribution: at a moment when there is a tie all possible guesses about the fraction of time during which one partly has been leading are equally probable. This result contrasts not only with Theorem 1 but even more with a result of Paul Lévy[3] (Corollary 2, pp. 303–304). According to Lévy, if the condition $S_{2n} = 0$ is replaced by the hypothesis that at the $2n$th trial $S_{2n}$ vanishes for the $k$th time, the limiting distribution as $k \to \infty$ of Theorem 2A is again the arc sin law ($a$) and no longer ($b$).

These results should serve as a warning to statisticians who might assume that fluctuation phenomena always follow the bell-shaped pattern and who would easily discover secular trends. If a coin is tossed once a second for a total of 365 days, the probability that one of the players will lead for more than 364 days and 10 hours is about 0.05! However, if it is known that the game concluded at a moment where neither player had a gain or loss, then the probability of such an extended lead is less than 0.0002.

*Proof of Theorem 1.* Let

$$p_{2r}(2n) = P(N_{2n} = 2r) \qquad \text{for } n > 0$$

$$p_0(0) = 1, \qquad p_r(0) = 0 \qquad \text{for } r > 0.$$

$$f_{2k} = P(S_j \neq 0 \text{ for } 1 \leq j < 2k;\ S_{2k} = 0) \qquad \text{for } k \leq 1;\ f_0 = 0$$

$$F(t) = \sum_{k=0}^{\infty} f_{2k}t^{2k} = 1 - (1 - t^2)^{1/2}.$$

Unless $r = 0$ or $r = n$, there is a smallest $k$ such that $1 \leq k < n$ and $S_{2k} = 0$. All $S_j$ with $1 \leq j \leq 2k$, are either all positive or all negative. These considerations lead to the recurrence relation

$$p_{2r}(2n) = \frac{1}{2}\sum_{k=0}^{n} f_{2k}p_{2r-2k}(2n - 2k) + \frac{1}{2}\sum_{k=0}^{n} f_{2k}p_{2r}(2n - 2k). \quad (1)$$

For $r = n$ we must add the probability that $S_{2k} > 0$ for $1 \leq k \leq 2n$, which is equal to

$$\frac{1}{2}\sum_{k=n+1}^{\infty} f_{2k} = \frac{1}{2}\binom{2n}{n}\frac{1}{2^{2n}}. \quad (2)$$

For $r = 0$ we must add the same quantity which now represents the probability that $S_{2k} < 0$ for $k \leq 2n$.

Introducing the generating function

$$g_{2n}(s) = \sum_{r=0}^{2n} p_{2r}(2n)s^{2r},$$

we obtain from formula (1)

$$g_{2n}(s) = \frac{1}{2}\sum_{k=0}^{n} f_{2k}s^{2k}g_{2n-2k}(s) + \frac{1}{2}\sum_{k=0}^{n} f_{2k}g_{2n-2k}(s) + \frac{1 + s^{2n}}{2}\binom{2n}{n}\frac{1}{2^{2n}}. \quad (3)$$

We now introduce the double generating function

$$G_s(t) = \sum_{i=0}^{\infty} g_{2i}(S)t^{2i}.$$

Then from formula (3),

$$G_s(t) = \frac{1}{2}\{G_s(t)F(st) + G_s(t)F(t)\} + \frac{1}{2}\{(1 - t)^{-1/2} + (1 - st)^{-1/2}\}, \quad (4)$$

where $F(t) = 1 - (1 - t^2)^{1/2}$, as implied by formula (2). Solving for $G_s(t)$ we get

$$G_s(t) = \frac{(1 - t)^{-1/2} + (1 - st)^{-1/2}}{(1 - t)^{1/2} + (1 - st)^{1/2}} = \frac{1}{(1 - t)^{1/2}(1 - st)^{1/2}}.$$

The coefficient of $s^{2r}t^{2n}$ is $p_{2r}(2n)$ and is that given in Theorem 1.

*Proof of Theorem 2.* We use the same notations as in the preceding proof except that now

$$p_{2r}(2n) = P(N_{2n} = 2r;\ S_{2n} = 2m).$$

The formula (1) remains valid except that for $r = n$ we must now add

$$W_n = P(S_k > 0 \text{ for } 0 < k \leq n;\ S_{2n} = 2m) = \frac{m}{n}\binom{2n}{n-m}\frac{1}{2^{2n}}$$

by a well-known formula (see reference 4, p. 153). For $r = 0$, formula (1) is valid as it stands with both sides vanishing. Using the generating function

$$W(s) = \sum_{n=m}^{\infty} w_{2n}s^{2n} = \left[\frac{1-(1-s^2)^{1/2}}{s}\right]^m$$

we obtain, instead of formula (4),

$$G_s(t) = {}^1/_2\{G_s(t)F(st) + G_s(t)F(t)\} + W(st).$$

Hence

$$G_s(t) = \frac{2W(st)}{(1-t^2)^{1/2} + (1-s^2t^2)^{1/2}} = 2W(st)\,\frac{(1-s^2t^2)^{1/2} - (1-t^2)^{1/2}}{t^2(1-s^2)}. \tag{5}$$

For $m = 0$, $W(s) \equiv 1$. The coefficient of $s^{2r}t^{2n}(r \leq n)$ in the right-hand side of formula (5) is that of $s^{2r}t^{2n+2}$ in

$$-2(1-t^2)^{1/2}(1-s^2)^{-1}.$$

Thus it is the same for all $r$, $0 \leq r \leq n$. This means that $p_{2r}(2n)$, and consequently also $P(N_{2n} = 2r|S_{2n} = 0)$, is the same for all $r$, $0 \leq r \leq n$. Hence it is equal to $\frac{1}{n+1}$. Theorem 2A is proved.

If $m > 0$, the same coefficient is easily seen to be

$$\sum_{i=m}^{r}\left|2\binom{1/2}{n-i+1}\right| w_{2i}.$$

This is equivalent to the formula given in Theorem 2.

[*] This paper is connected with work under an ONR contract for research in probability theory at Cornell University.

[1] Erdös, P., and Kac, M., "On the Number of Positive Sums of Independent Random Variables," *Bull. Am. Math. Soc.*, **53**, 1011–1020 (1947).

[2] Andersen, E. S., "On the Number of Positive Sums of Random Variables," *Skand. Aktuarietid*, 1950, to appear.

[3] Lévy, P., "Sur certains processus stochastiques homogènes," *Compositio Math.*, **7**, 283–339 (1939).

[4] Uspensky, V., *Introduction to Mathematical Probability*, McGraw-Hill, New York, 1937.

# *A GENERALIZATION OF THE SCHWARZ-CHRISTOFFEL TRANSFORMATION**

By D. Gilbarg

Indiana University

Communicated by T. Y. Thomas, August 30, 1949

The classical Schwarz-Christoffel transformation provides a formula for the conformal mapping of the half-plane onto a plane polygon. A generalization enlarging the class of image domains to include many-sheeted plane polygons with interior winding points was known already to at least Schwarz and Schläfli.[1] This note is concerned with a further simple extension of the formula which delivers a useful class of mappings defined by multiple-valued functions. These mappings have been applied elsewhere by the author to certain hydrodynamical problems.[2]

Let $w(z)$ be a multiple-valued analytic function defined on the half-plane, $Im\ z > 0$. It is assumed that any regular function element of $w(z)$ can be continued along arbitrary paths to all but a finite number of points. It will be convenient in what follows to consider the image of a single-valued branch of $w(z)$ obtained by making a suitable cut or cuts in the half-plane which connect all the branch points of $w(z)$ with the boundary. If, on the image Riemann surface thus defined over the $w$-plane, the images of opposite points on the two edges of the cut are identified, we obtain an *ideal Riemann surface* or *Riemannian manifold* in the sense of Koebe.[3] (The image of a branch point, if it exists, is included as interior point of the surface.) This domain is a single-valued image of the half-plane under the mapping defined by (a branch of) $w(z)$. In the following we consider the conformal mapping of the half-plane onto certain Riemannian manifolds with polygonal boundaries. We have first,

Theorem 1. *Let $w(z)$ be a possibly multiple-valued analytic function defined on the half-plane, $Im\ z > 0$. Let the function elements for $w(z)$ be such that in the neighborhood of any point $z = a$,*

$$w'(z) = (z-a)^{\alpha} P(z-a),\ P(0) \neq 0, \tag{1}$$

*where $P(z-a)$ is regular at $z = a$, and $\alpha$ is non-zero for at most a finite number of points, $z = a_j$. Let $w(z)$ map the half-plane onto a Riemannian manifold $M$ with a closed polygonal boundary having interior angles $\beta_k\pi$ at vertices which are images by $w(z)$ of the points $z = b_k (b_k$ real). Then $w(z)$ is of the form,*

$$w(z) = A \int^{z} \{\prod_j [(z - a_j)^{\alpha_j} (z - \bar{a}_j)^{\bar{\alpha}_j}] \prod_k (z - b_k)^{\beta_k - 1}\}\, dz + B, \tag{2}$$

*where $A$ and $B$ are complex constants.*

*Proof:* The proof follows basically that of the classical Schwarz-Christoffel formula. We observe first that any two function elements for $w(z)$ over the same point are linearly related, by virtue of a standard argument based on equation (1). Therefore the image of the real axis under the mapping determined by any branch of $w(z)$ is polygonal, the vertices being images of the points $z = b_k$, and $\beta_k\pi$ the magnitude of the corresponding angles. Hence $w(z)$ (whatever the branch) can be continued analytically into the lower half plane by reflection across the real axis. On the real axis $w(z)$ satisfies equation (1), the exponent $\alpha$ being equal to $\beta_k - 1$ at $z = b_k$ since the straight angle at $b_k$ maps into an angle of magnitude $\beta_k\pi$; $\alpha$ is zero elsewhere on the real axis. Consider now the function $w''/w'$. By virtue of equation (1) and the foregoing, this function must be regular in the finite plane except for simple poles of residue $\alpha_j$, $\bar{\alpha}_j$, $\beta_k - 1$, at $z = a_j$, $\bar{a}_j$, $b_k$, respectively. At the point at infinity, which is mapped by $w(z)$ into a boundary point, $w''/w'$ must be regular and also vanishes. It follows that $w''/w'$ is single-valued in the full $z$-plane, and regular except for the above-mentioned simple poles. We therefore have,

$$\frac{w''}{w'} = \frac{\alpha_j}{z - a_j} + \frac{\bar{\alpha}_j}{z - \bar{a}_j} + \frac{\beta_k - 1}{z - b_k},$$

from which equation (2) is obtained by integration.

If the exponent $\alpha$ in equation (1) is zero for all interior points, then equation (2) reduces to the classical Schwarz-Christoffel formula. Under certain circumstances, for example, if all the non-zero exponents $\alpha_j$ are positive integers, the (single-valued) mapping in equation (2) sends the half-plane into a many-sheeted Riemann surface with polygonal boundary, the images of the $a_j$ being winding points of the surface.

The most general algebraic and logarithmic singularities are obviously not included in equation (1), and whether a generalization of equation (2) can be obtained for functions having such singularities is an open question.

Any Riemannian manifold $M$ which is image of the half-plane under the mapping, equation (2), evidently has a polygonal boundary, is simply connected and contains a finite number of algebraic winding points. In addition, the coördinating transformations (Koebe: Bezugstransformation) which identify arcs on $M$ are linear. For, it follows from equation (2) that if $C$ is any arc in the $z$-plane containing no branch points of $w(z)$ in its interior, and $C^+$, $C^-$ two arcs in the $w$-plane which are images of opposite edges of $C$, then a linear transformation of the $w$-plane transforms $C^-$ into $C^+$. The vertices of the coördinating transformations (namely, the images of the branch points), are finite in number.

Conversely, let $M$ be a Riemannian manifold with the above properties. Then by the principal theorems on the conformal mapping of Riemannian manifolds,[3] there is a function $z(w)$ which gives a $1-1$ mapping of $M$ onto the open half-plane, any pair of coördinated arcs on $M$ mapping into the same arc in the $z$-plane. The latter arcs, considered as cuts in the $z$-plane, define the branch of the inverse function $w(z)$ having $M$ as image. Let $w_+(z)$, $w_-(z)$ be function elements for $w(z)$ on opposite sides of an arc $C$ of such a cut, assuming for convenience that $C$ contains no branch points of $w(z)$. Consider the analytic continuations of $w_+(z)$ and $w_-(z)$ across $C$; (that such continuation is always possible follows from the definition of Riemannian manifold). Then, since $w_+(z)$ and $w_-(z)$ are linearly related on $C$, this relation holds in a full neighborhood of $C$. It follows that the functions $w_+''/w_+'$ and $w_-''/w_-'$ are identical in this neighborhood, and, by a simple consideration, that $w''/w'$ is single-valued in the entire half plane. Hence, by using the Laurent expansion for $w''/w'$, the function elements for $w(z)$ are seen to satisfy at any point $a$.

$$w'(z) = (z-a)^{\alpha} P(z-a) \exp Q(1/z-a),$$

where $P(z-a)$, $Q(1/z-a)$ are, respectively, power series in $z-a$ and $1/(z-a)$. However, the exponent $Q(1/z-a)$ must be constant, for otherwise, $\lim\limits_{z \to a} w(z)$ would not exist and define an interior point of $M$, as required. Thus, the function elements of $w(z)$ satisfy equation (1); and since the number of winding points and vertices (of coördinations) on $M$ is by assumption finite, the conditions on $w(z)$ for theorem 1 apply. Therefore, we have,

THEOREM 2. *Let $M$ be a simply connected Riemannian manifold over the $w$-plane with polygonal boundary. Let the coördinating transformations on $M$ be linear, the number of vertices of the transformations being finite. All other points on $M$ are to be regular except for a finite number of algebraic winding points. Then there is a function $w(z)$ which maps the half-plane, $Im\ z > 0$, onto $M$, where $w(z)$ is of the form,*

$$w(z) = A \int^{z} \{\prod_{j}[(z - a_j)^{\alpha_j} (z - \bar{a}_j)^{\bar{\alpha}_j}] \prod_{k}(z - b_k)^{\beta_k - 1}\} \, dz + B.$$

*In this expression, $\beta_k\pi$ are the interior angles of the polygonal boundary of $M$, and $A$, $B$, $a_j$, $b_k$, $\alpha_j$, are constants, with $Im\ a_j > 0$, $b_k$ real.*

The actual determination of the constants in equation (2) is a matter of great practical difficulty, as it is for the classical Schwarz-Christoffel formula. The preceding theorem establishes only the existence of appropriate constants so that a branch of the function defined by equation (2) gives the desired mapping. Since a transformation of the half-plane into itself preserves the mapping, three of the constants $b_k$ may be preassigned, or one $a_j$ and $b_k$.

---

* Prepared under Navy Contract No. N6onr-180, Task Order 5, with Indiana University.

[1] Schwarz, H. A., *Ges. Math. Abh*, II, 77 (1890); Schläfli, L., *J. reine angew. Math.*, 78, 63–80 (1874).

[2] Gilbarg, D., and Serrin, J., *Free Boundaries and Jets in the Theory of Cavitation*, forthcoming publication.

[3] Koebe, P., "Allgemeine Theorie der Riemannschen Mannigfaltigkeiten," *Acta Math.*, 50, 27–157 (1927). See also Courant, R., *Dirichlet's Principle*, Interscience (to be published); this uses the terminology *Riemann domain*.

---

# *ON SOME GENERALIZATIONS OF THE CAUCHY-FRULLANI INTEGRAL**

By A. M. Ostrowski

University of Basle, Switzerland; U. S. National Bureau of Standards; and University of California at Los Angeles

Communicated by E. U. Condon, July 28, 1949

1. A beautiful result attributed often to Frullani is contained in the general formula

$$\int_0^\infty \frac{f(at) - f(bt)}{t}\,dt = (f(\infty) - f(0)) \log \frac{a}{b}\ (a, b > 0), \qquad (1)$$

where $f(0) = \lim\limits_{x \downarrow 0} f(x)$, $f(\infty) = \lim\limits_{x \to \infty} f(x)$ and $f(x)$ is assumed $L$-integrable over any interval $0 < A \leq x \leq B < \infty$.[1]

2. Suppose that the integral

$$\int_0^A \frac{f(t)}{t}\,dt \qquad (2)$$

exists for any $A > 0$. Then the above formula can be replaced by

$$\int_0^\infty \frac{f(at) - f(bt)}{t}\,dt = f(\infty) \log \frac{a}{b}\ (a, b > 0) \qquad (3)$$

if $f(\infty) = \lim\limits_{x \to \infty} f(x)$ exists. It can be expected that this formula remains valid if $f(\infty)$ is replaced by appropriate *mean values*.

Suppose, for instance, that $f(x)$ is *periodic with period $p$* and the integral (2) exists for any $A > 0$. Then $f(\infty)$ can be replaced by $\frac{1}{p}\int_u^{u+p} f(x)dx$ and we get the following very useful formula

$$\int_0^\infty \frac{f(at) - f(bt)}{t}\,dt = \frac{1}{p}\int_0^{u+p} f(x)dx \log \frac{a}{b}. \qquad (4)$$

3. Thus we obtain from

$$\int_0^\pi |\tan x|^\alpha dx = \frac{\pi}{\cos\frac{\alpha\pi}{2}} \quad (0 < \alpha < 1),$$

the formula

$$\int_0^\infty (|\tan ax|^\alpha - |\tan bx|^\alpha)\frac{dx}{x} = \frac{1}{\cos\frac{\alpha\pi}{2}}\log\frac{a}{b} \quad (0 < \alpha < 1;\ a, b > 0). \tag{5}$$

Similarly, it follows from

$$\int_0^\pi \log|\cos x|dx = \pi \log {}^1/_2$$

that

$$\int_0^\infty \log\left|\frac{\cos ax}{\cos bx}\right|\frac{dx}{x} = \log {}^1/_2 \log\frac{a}{b}\ (a, b > 0). \tag{6}$$

4. The condition of the periodicity of $f(t)$ can be replaced by an essentially more general one, that of the existence of the mean value

$$M(f) = \lim_{x\to\infty}\frac{1}{x}\int_0^x f(t)dt \tag{7}$$

and we obtain the following general theorem:

*If the integral* (2) *exists for any* $A > 0$ *and the mean value* (7) *exists, we have for all positive* $a$ *and* $b$

$$\int_0^\infty \frac{f(at) - f(bt)}{t}\, dt = M(f)\log\frac{a}{b}. \tag{8}$$

5. To prove (8) we use the fact that

$$F(x) = \frac{1}{x}\int_0^x f(t)dt \qquad (x > 0)$$

is continuous for all $x > 0$ and tends to $M(f)$ as $x \to \infty$. We have on integrating by parts

$$\int_{bA}^{aA}\frac{f(t)}{t}\,dt = \int_{bA}^{aA}\frac{(tF(t))'}{t}\,dt = F(aA) - F(bA) + \int_{bA}^{aA}\frac{F(t)}{t}\,dt = F(aA) - F(bA) + F(\zeta)\log\frac{a}{b}$$

where $bA < \zeta < aA$. It follows

$$\lim_{A\to\infty}\int_{bA}^{aA}\frac{f(t)}{t}\,dt = M(f)\log\frac{a}{b}.$$

On the other hand we have

$$\int_0^{aA}\frac{f(t)}{t}\,dt - \int_0^{bA}\frac{f(t)}{t}\,dt = \int_0^{A}\frac{f(at)}{t}\,dt - \int_0^{A}\frac{f(bt)}{t}\,dt = \int_0^{A}\frac{f(at)-f(bt)}{t}\,dt$$

and our theorem is proved.

6. The assumption of the existence of (2) can be decomposed into two parts: (1) $f(t)$ is $L$-integrable over any interval $0 < \epsilon \leq t \leq A < \infty$; (2) $\frac{f(t)}{t}$ is $L$-integrable over any interval $0 \leq t \leq \epsilon$. The second part can be replaced by the assumption that $\lim\limits_{\epsilon \downarrow 0} \int_\epsilon^A \frac{f(t)}{t}\,dt$ exists or by the hypothesis that $f(0) = \lim\limits_{\epsilon \downarrow 0} f(t)$ exists and has the value 0. Both hypotheses are only special cases of a more general one which is obtained from the assumption about the behavior of $f(t)$ at $t = \infty$ by the transformation $t = \frac{1}{\tau}$. We come thus to the assumption

$$\epsilon \int_\epsilon^1 \frac{f(t)}{t^2}\,dt \to 0 \qquad (\epsilon \downarrow 0).$$

More generally if we assume that together with (7)

$$m(f) = \lim_{\epsilon \downarrow 0} \epsilon \int_\epsilon^1 \frac{f(t)}{t^2}\,dt \tag{9}$$

exists, we have

$$\int_0^\infty \frac{f(at)-f(bt)}{t}\,dt = (M(f) - m(f)) \log \frac{a}{b}. \tag{10}$$

7. The result obtained in (10) is in a certain sense the best result obtainable. It can be shown that if the integral at the left in (10) exists for some pairs of positive values $a$, $b$ such that $\frac{b}{a}$ runs through a set of positive measure, then both limits $M(f)$, $m(f)$ exist and we have (10). However, the proof of this result is difficult and will be given in another publication.

Since finding the formula (10) I have discovered that the problem of convergence of the integral at the left in (10) has already been investigated and solved by K. S. K. Iyengar.[7] The necessary and sufficient conditions given by Iyengar consist in the existence of the four following limits:

$$\lim_{x \to \infty} \int_1^x \frac{f(t)}{t^2}\,dt, \qquad \lim_{x \to \infty} x \int_x^\infty \frac{f(t)}{t^2}\,dt \tag{11}$$

$$\lim_{x\downarrow 0}\int_x^1 f(t)dt, \qquad \lim_{x\downarrow 0}\frac{1}{x}\int_0^x f(t)dt. \tag{12}$$

It can be directly proved that both conditions (11) are equivalent to the existence of $M(f)$. Similarly, both conditions (12) are equivalent to the existence of $m(f)$. Iyengar's proof of his theorem and the simplified proof of it given by Agnew[2] are, however, still difficult since in both proofs certain theorems about non-uniform convergence are used which, although usually proved only in the case of convergent *sequences*, have to be used in the case of *continuous approximation*. Our proof of the necessity of the existence of $M(f)$ and $m(f)$ makes use of the theorem of Osgood, but exactly in the form proved by Osgood, that is for the case of *sequences* of functions.

8. Formula (1) can be considerably generalized by replacing $at$ and $bt$ by functions of $t$ which to a certain extent are arbitrary. In this way we obtain a "three-function formula" giving the value of definite integrals containing three "arbitrary" functions.

This general formula cannot be directly reduced to (1) by substituting a new variable of integration. However, the interval of integration and the range of values of two of the functions can be reduced by such a transformation to the interval $(0, \infty)$. We obtain thus the canonical form of our formula

$$\int_0^\infty (\psi' g(\psi) - \varphi' g(\varphi))dx = M(xg(x)) \log\left(\frac{\psi}{\varphi}\right)_{x=\infty} - m(xg(x)) \log\left(\frac{\psi}{\varphi}\right)_{x=0}. \tag{13}$$

The functions $\varphi(x)$ and $\psi(x)$ are supposed in (13) to be positive and absolutely continuous for $0 < x < \infty$ and to tend to 0 with $x \downarrow 0$ and to $\infty$ with $x \to \infty$. We assume further that with $x \downarrow 0$ and $x \to \infty$ the limits of $\frac{\varphi(x)}{\psi(x)}$ exist and are positive. The function $g(x)$ is assumed to be integrable over any closed interval of the positive $x$-axis and such that $M(xg(x))$ and $m(xg(x))$ exist.

9. Three special cases of (13) have been previously given. In 1823 Cauchy[3] gave the formula

$$\int_0^1 \left(\frac{\psi' f(\psi)}{1-\psi} - \frac{\varphi' f(\varphi)}{1-\varphi}\right) dx = f(1) \log \frac{\varphi'(1)}{\psi'(1)}. \tag{14}$$

Here $f(x)$ is continuous in the closed interval $< 0, 1 >$; $\varphi$ and $\psi$ are differentiable in $< 0, 1 >$ with positive derivatives and it is assumed that $\varphi(0) = \psi(0) = 0$, $\varphi(1) = \psi(1) = 1$. In 1841 Cauchy[4] published another formula

$$\int_0^\infty \left(\psi' \frac{f(\psi)}{\psi} - \varphi' \frac{f(\varphi)}{\varphi}\right) dx = f(0) \log \frac{\varphi'(0)}{\psi'(0)}. \tag{15}$$

In this formula $\varphi$ and $\psi$ are positive in $(0, \infty)$ and have continuous derivatives which are positive at 0. Further we assume that $\varphi(0) = \psi(0) = 0$ and that as $x \to \infty$ $\varphi(x)$ and $\psi(x)$ tend to $\infty$. About $f(x)$ Cauchy assumes continuity for all $x \geq 0$. However, about the behavior of $f(x)$ for $x \to \infty$ Cauchy erroneously gives the condition $f(x) \to 0$. This condition is not sufficient for the validity of the above formula but Cauchy's argument remains valid if the existence of $\int_1^\infty \frac{f(x)}{x}\, dx$ is assumed.

Finally, 1891, M. Lerch[6] gave the formula

$$\int_a^b (g(x) - \varphi' g(\varphi(x)))dx = \\ ((x - a)g(x))_{x=a} \log \varphi'(a) - ((x - b)g(x))_{x=b} \log \varphi'(b) \tag{16}$$

that is obtained from (13) by obvious specializations and transformations.

Some of the results given in this note have been discovered independently by Professor Tricomi of the California Institute of Technology in his work on the Bateman manuscript project and are going to be published in the *Bulletin of the American Mathematical Society*. Professor Tricomi discovered our formula (4); further, a case of our formula (8), namely the case in which

$$\int_0^x f(x)dx - xM(f)$$

is $O(1)$—while in our theorem it is assumed that this difference is $o(x)$; as to the three-function formula, he finds the special case in which $g(x)$ has finite limits at 0 and $\infty$, respectively, and $xg(x)$ is periodic.

---

* The preparation of this paper was sponsored (in part) by the Office of Naval Research.

[1] This formula was first published by Cauchy in 1823,[3] and more completely 1827[3a] with a beautiful proof which is used as standard in all textbooks today. About 1829 Frullani[5] published the same formula and mentioned that he had communicated it to plana in 1821. However, Frullani's proof is completely illusory.

[2] Agnew, R. P., "Limits of Integrals," *Duke Math. J.*, **9**, 10–19 (1942).

[3] Cauchy, A., *J. École Polytech. (Paris)*, **12** (1823); *Oeuvres compl.* (2), **I**, 335–339.

[3a] Cauchy, A., *Exercises de Mathematiques*, 1827; *Oeuvres compl.* (2), **VII**, 157.

[4] Cauchy, A., *Exercises Analyse*, **2** (1841); *Oeuvres compl.* (2), **XII**, 416–417.

[5] Frullani, G., "Sopra Gli Integrali Definiti (Ricevuta adi 21 Novembre (1829)," *Memorie della Societa Italiana delle Socienze*, **20**, 448–467 (1828).

[6] Lerch, M., "Sur une extension de la formule de Frullani," *Verhandl. Prager Akad., math.-phys. Klasse* $I_2$, 123–131 (1891).

[7] Iyengar, K. S. K., "On Frullani Integrals," *J. Indian Math. Soc.* (2), **4**, 145–150 (1940); reprinted in *Proc. Cambridge Phil. Soc.*, **37**, 9–13 (1941).

PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES

Volume 35 November 15, 1949 Number 11

# FIRST APPROXIMATION OF PRESSURE DISTRIBUTION ON CURVED PROFILES AT SUPERSONIC SPEEDS

By T. Y. Thomas

Department of Mathematics, University of Indiana

Communicated September 14, 1949

1. *Introduction.*—In a paper[1] recently written we considered the problem of pressure distribution on curved profiles in a field of uniform supersonic gas flow under the assumption that a shock wave originated at the vertex of the obstacle. Behind the shock the flow was assumed non-isentropic and rotational, although viscosity and thermal conductivity were neglected. The pressure $p$ was represented by a function $p(\omega, S)$ where $\omega$ denotes the inclination of the stream lines and $S$ the entropy. Choosing the arbitrary constant in the entropy function so that $S = 0$ along the profile the required pressure distribution is given by $\theta(\omega) = p(\omega, 0)$ in which $\omega$ is the inclination at points of the profile. A formal solution of the flow problem is obtained which satisfies the boundary and shock conditions, provided that the function $\theta(\omega)$ satisfies an ordinary differential equation of infinite order. This differential equation appears in the form of a power series with coefficients which are functions of the inclination $\omega$ of the profile; moreover the coefficients, after the first, involve derivatives of the function $\theta(\omega)$. When this power series is terminated after the sum of its first $n$ terms, we obtain an equation for the $n$th approximation of pressure on the boundary.

It is the purpose of this paper[2] to treat the determination of pressure by means of the first approximation which is given by

$$\frac{\theta}{p_1} = \frac{2\gamma}{\gamma + 1}(M^2 \sin^2 \alpha - 1) + 1, \qquad (1)$$

where $M$ is the Mach Number of the uniform flow before the shock, $p_1$ its pressure, and the inclination $\omega$ of the profile is related to the parameter $\alpha$ in (1) by the equation

$$\tan \omega = \frac{2[(M^2 - 1)\tan^2 \alpha - 1]}{[(\gamma - 1)M^2 + 2]\tan^3 \alpha + [(\gamma + 1)M^2 + 2]\tan \alpha} \quad (2)$$

We take the constant $\gamma$ to have the usual value 1.405. *The pressure given by this approximation at a point of the profile at which the inclination is* $\omega$ *is the same as that immediately behind the shock line at the point where* $\omega$ *gives the inclination of the stream line.*

2. *Computations and Graphs.*—Consider the profile formed by two circular arcs of radius $r = 150$ and chord length 60 as shown in figure 1. We shall calculate the pressure distribution on the upper contour of this profile for various values of the angle of attack $a$ when the incident flow has Mach Number $M = 2.13$. This particular profile and Mach Number are selected, since these have been used by A. Ferri[8] in connection with his pressure determinations at the high-speed tunnel at Guidonia. Thus we shall be in a position to compare our calculated values of pressure with experimental determinations.

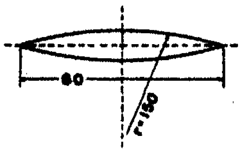


FIGURE 1

Table 1 shows the relation between $\alpha$ and $\omega$ computed in accordance with equation (2) for values of $\alpha$ ranging between 10.2667 (for which $\theta/p_1$ vanishes) and 65, inclusive. By interpolation from this table, we can find the value of $\alpha$ corresponding to any point of the profile at which $\omega$ is known and hence by means of (1) we can find the value of the pressure ratio $\theta/p_1$ at this point. At the per cent chord of the profile having the value $\beta$ it is easily seen that the value of $\omega$ is given by

$$\sin \omega = 0.2 - \beta/250.$$

Using this relation, we can find the value of $\omega$ at the per cent chord $\beta$ and hence the values of $\alpha$ and $\theta/p_1$ as above stated.

In tabulating his results Ferri has referred the values of the pressure along the profile to the value of the pressure before the shock, and for purposes of comparison we shall adopt the same procedure. Thus, instead of the ratio $\theta/p_1$ we shall list the values of the ratio

$$P = \frac{\theta - p_1}{p_1}.$$

At a point of the profile at which the pressure is greater than $p_1$ the value of $P$ will be positive and the value of $P$ will be negative whenever the pressure is less than the pressure in front of the shock.

The results of these calculations are entered in tables 2 to 11. Each table corresponds to a different angle of attack $a$ which is given at the top of the table. Also the experimental value of the pressure $p_1$ is shown at

TABLE 1

| $\alpha$(DEG.) | $\omega$(DEG.) | $\theta/p_1$ | $P$ | $\alpha$(DEG.) | $\omega$(DEG.) | $\theta/p_1$ | $P$ |
|---|---|---|---|---|---|---|---|
| 10.2667 | −36.8330 | 0.0000 | −1.0000 | 38 | 11.0110 | 1.8411 | 0.8411 |
| 11 | −34.3208 | 0.0245 | −0.9755 | 39 | 11.9278 | 1.9307 | 0.9307 |
| 12 | −31.1333 | 0.0606 | −0.9394 | 40 | 12.8167 | 2.0219 | 1.0219 |
| 13 | −28.1875 | 0.0999 | −0.9001 | 41 | 13.6777 | 2.1136 | 1.1136 |
| 14 | −25.4500 | 0.1417 | −0.8583 | 42 | 14.5167 | 2.2049 | 1.2049 |
| 15 | −22.9000 | 0.1868 | −0.8132 | 43 | 15.3167 | 2.2970 | 1.2970 |
| 16 | −20.5112 | 0.2345 | −0.7655 | 44 | 16.1042 | 2.3898 | 1.3898 |
| 17 | −18.2612 | 0.2848 | −0.7152 | 45 | 16.8500 | 2.4821 | 1.4821 |
| 18 | −16.1500 | 0.3379 | −0.6621 | 46 | 17.5777 | 2.5743 | 1.5743 |
| 19 | −14.1722 | 0.3935 | −0.6065 | 47 | 18.2722 | 2.6671 | 1.6671 |
| 20 | −12.2778 | 0.4518 | −0.5482 | 48 | 18.9417 | 2.7588 | 1.7588 |
| 21 | −10.4833 | 0.5128 | −0.4872 | 49 | 19.5792 | 2.8510 | 1.8510 |
| 22 | − 8.7888 | 0.5753 | −0.4247 | 50 | 20.1888 | 2.9422 | 1.9422 |
| 23 | − 7.1612 | 0.6405 | −0.3595 | 51 | 20.7722 | 3.0328 | 2.0328 |
| 24 | − 5.6083 | 0.7084 | −0.2916 | 52 | 21.3167 | 3.1230 | 2.1230 |
| 25 | − 4.1222 | 0.7784 | −0.2216 | 53 | 21.8333 | 3.2125 | 2.2125 |
| 26 | − 2.6888 | 0.8504 | −0.1496 | 54 | 22.3167 | 3.3010 | 2.3010 |
| 27 | − 1.3222 | 0.9242 | −0.0758 | 55 | 22.7625 | 3.3890 | 2.3890 |
| 28 | − 0.0000 | 0.9999 | −0.0001 | 56 | 23.1750 | 3.4744 | 2.4744 |
| 29 | 1.2722 | 1.0773 | 0.0773 | 57 | 23.5500 | 3.5603 | 2.5603 |
| 30 | 2.5000 | 1.1568 | 0.1568 | 58 | 23.8888 | 3.6435 | 2.6435 |
| 31 | 3.6888 | 1.2374 | 0.2374 | 59 | 24.1833 | 3.7267 | 2.7267 |
| 32 | 4.8388 | 1.3201 | 0.3201 | 60 | 24.4375 | 3.8073 | 2.8073 |
| 33 | 5.9500 | 1.4038 | 0.4038 | 61 | 24.6458 | 3.8863 | 2.8863 |
| 34 | 7.0278 | 1.4892 | 0.4892 | 62 | 24.8055 | 3.9636 | 2.9636 |
| 35 | 8.0667 | 1.5756 | 0.5756 | 63 | 24.9167 | 4.0399 | 3.0399 |
| 36 | 9.0778 | 1.6631 | 0.6631 | 64 | 24.9750 | 4.1137 | 3.1137 |
| 37 | 10.0612 | 1.7516 | 0.7516 | 65 | 24.9792 | 4.1858 | 3.1858 |

TABLE 2

$a = 0$; $p_1 = 52.1$; $q = 165.5$

| % CHORD | $\omega$(DEG.) | $\alpha$(DEG.) | CAL. $P$ | EXP. $P$ | CAL. $Q$ | EXP. $Q$ |
|---|---|---|---|---|---|---|
| 0.0 | 11.5388 | 38.5757 | 0.8927 | .... | 4.2014 | .... |
| 4.0 | 10.6000 | 37.5673 | 0.8024 | 0.8253 | 3.7764 | 3.8973 |
| 12.5 | 8.6278 | 35.5549 | 0.6242 | 0.5566 | 2.9377 | 2.6284 |
| 21.5 | 6.5445 | 33.5516 | 0.4509 | 0.3839 | 2.1221 | 1.8127 |
| 31.0 | 4.3612 | 31.5847 | 0.2858 | 0.2399 | 1.3451 | 1.1329 |
| 40.5 | 2.1778 | 29.7376 | 0.1359 | 0.0960 | 0.6396 | 0.4532 |
| 50.5 | − 0.1167 | 27.9117 | −0.0068 | −0.0289 | −0.0320 | −0.1360 |
| 59.0 | − 2.0612 | 26.4592 | −0.1157 | −0.1440 | −0.5445 | −0.6798 |
| 68.5 | − 4.2445 | 24.9177 | −0.2274 | −0.2591 | −1.0702 | −1.2236 |
| 78.5 | − 6.5945 | 23.3649 | −0.3347 | −0.3551 | −1.5752 | −1.6767 |
| 87.5 | − 8.6722 | 22.0716 | −0.4200 | −0.2975 | −1.9767 | −1.4048 |
| 97.0 | −10.8333 | 20.8050 | −0.4991 | −0.2975 | −2.3490 | −1.4048 |

TABLE 3

$a = -2$; $p_1 = 52.2$; $q = 166$

| % CHORD | $\omega$(DEG.) | $\alpha$(DEG.) | CAL. $P$ | EXP. $P$ | CAL. $Q$ | EXP. $Q$ |
|---|---|---|---|---|---|---|
| 0.0 | 13.5388 | 40.8387 | 1.0988 | .... | 5.1714 | .... |
| 4.0 | 12.6000 | 39.7562 | 0.9997 | 0.9962 | 4.7050 | 4.6988 |
| 12.5 | 10.6278 | 37.5965 | 0.8050 | 0.7471 | 3.7887 | 3.5241 |
| 21.5 | 8.5445 | 35.4726 | 0.6170 | 0.7184* | 2.9038 | 3.3886* |
| 31.0 | 6.3612 | 33.3815 | 0.4364 | 0.4215 | 2.0539 | 1.9880 |
| 40.5 | 4.1778 | 31.4252 | 0.2726 | 0.2874 | 1.2830 | 1.3554 |
| 50.5 | 1.8333 | 29.5611 | 0.1219 | 0.2107 | 0.5737 | 0.9940 |
| 59.0 | −0.0612 | 27.9537 | −0.0036 | 0.0575 | −0.0169 | 0.2711 |
| 68.5 | −2.2445 | 26.3251 | −0.1256 | 0.0000 | −0.5911 | 0.0000 |
| 78.5 | −4.5945 | 24.6822 | −0.2438 | −0.0766 | −1.1474 | −0.3614 |
| 87.5 | −6.6722 | 23.3149 | −0.3381 | −0.1628 | −1.5912 | −0.7681 |
| 97.0 | −8.8333 | 21.9737 | −0.4265 | −0.1916 | −2.0073 | −0.9036 |

* The figure in the table of experimental results given by Ferri, from which these values are derived, appears to be a misprint. We have therefore not indicated the corresponding experimental point on the graph in figure 3.

TABLE 4

$a = +4$; $p_1 = 52.3$; $q = 166.5$

| % CHORD | $\omega$(DEG.) | $\alpha$(DEG.) | CAL. $P$ | EXP. $P$ | CAL. $Q$ | EXP. $Q$ |
|---|---|---|---|---|---|---|
| 0.0 | 7.5388 | 34.4919 | 0.5317 | .... | 2.5024 | .... |
| 4.0 | 6.6000 | 33.6031 | 0.4553 | 0.4685 | 2.1428 | 2.2072 |
| 12.5 | 4.6278 | 31.8165 | 0.3049 | 0.2581 | 1.4350 | 1.2162 |
| 21.5 | 2.5445 | 30.0374 | 0.1598 | 0.1052 | 0.7521 | 0.4955 |
| 31.0 | 0.3612 | 28.2839 | +0.0221 | −0.0382 | +0.1040 | −0.1802 |
| 40.5 | − 1.8222 | 26.6341 | −0.1028 | −0.1052 | −0.4838 | −0.4955 |
| 50.5 | − 4.1167 | 25.0038 | −0.2213 | −0.2103 | −1.0415 | −0.9910 |
| 59.0 | − 6.0612 | 23.7084 | −0.3114 | −0.3250 | −1.4656 | −1.5315 |
| 68.5 | − 8.2445 | 22.3344 | −0.4029 | −0.3824 | −1.8962 | −1.8018 |
| 78.5 | −10.5945 | 20.9380 | −0.4910 | −0.3174 | −2.3108 | −1.4955 |
| 87.5 | −12.6722 | 19.7918 | −0.5603 | −0.3059 | −2.6370 | −1.4414 |
| 97.0 | −14.8333 | 18.6657 | −0.6251 | −0.3059 | −2.9420 | −1.4414 |

TABLE 5

$a = -6$; $p_1 = 52.3$; $q = 166.5$

| % CHORD | $\omega$(DEG.) | $\alpha$(DEG.) | CAL. $P$ | EXP. $P$ | CAL. $Q$ | EXP. $Q$ |
|---|---|---|---|---|---|---|
| 0.0 | 17.5388 | 45.9465 | 1.5699 | .... | 7.3886 | .... |
| 4.0 | 16.6000 | 44.3352 | 1.4207 | 1.4340 | 6.6864 | 6.7568 |
| 12.5 | 14.6278 | 42.1389 | 1.2177 | 1.1090 | 5.7310 | 5.2252 |
| 21.5 | 12.5445 | 39.6938 | 0.9940 | 0.8987 | 4.6782 | 4.2342 |
| 31.0 | 10.3612 | 37.3157 | 0.7799 | 0.7075 | 3.6705 | 3.3333 |
| 40.5 | 8.1778 | 35.1099 | 0.5852 | 0.5545 | 2.7542 | 2.6126 |
| 50.5 | 5.8333 | 32.8950 | 0.3950 | 0.4015 | 1.8590 | 1.8919 |
| 59.0 | 3.9399 | 30.2112 | 0.1738 | 0.2486 | 0.8180 | 1.1712 |
| 68.5 | 1.7555 | 29.3936 | 0.1086 | 0.1338 | 0.5111 | 0.6306 |
| 78.5 | −0.5945 | 27.4496 | −0.0418 | 0.0382 | −0.1967 | 0.1802 |
| 87.5 | −2.6722 | 26.0121 | −0.1487 | −0.0765 | −0.6998 | −0.3604 |
| 97.0 | −4.8333 | 24.5215 | −0.2551 | −0.1912 | −1.2006 | −0.9009 |

TABLE 6

$a = +8;\ p_1 = 52.2;\ q = 166.5$

| % CHORD | $\omega$(DEG.) | $\alpha$(DEG.) | CAL. $P$ | EXP. $P$ | CAL. $Q$ | EXP. $Q$ |
|---|---|---|---|---|---|---|
| 0.0 | 3.5388 | 30.8738 | 0.2272 | .... | 1.0693 | .... |
| 4.0 | 2.6000 | 30.0841 | 0.1636 | 0.1534 | 0.7700 | 0.7207 |
| 12.5 | 0.6278 | 28.4935 | +0.0381 | −0.0192 | 0.1793 | −0.0901 |
| 21.5 | − 1.4555 | 26.9025 | −0.0830 | −0.1246 | −0.3906 | −0.5856 |
| 31.0 | − 3.6388 | 25.3372 | −0.1973 | −0.2682 | −0.9286 | −1.2613 |
| 40.5 | − 5.8222 | 23.8623 | −0.3028 | −0.3448 | −1.4251 | −1.6216 |
| 50.5 | − 8.1167 | 22.4129 | −0.3978 | −0.4023 | −1.8722 | −1.8919 |
| 59.0 | −10.0612 | 21.2492 | −0.4716 | −0.4598 | −2.2195 | −2.1622 |
| 68.5 | −12.2445 | 20.0186 | −0.5471 | −0.4406 | −2.5749 | −2.0721 |
| 78.5 | −14.5945 | 18.7822 | −0.6186 | −0.4023 | −2.9114 | −1.8919 |
| 87.5 | −16.6722 | 17.7527 | −0.6752 | −0.4023 | −3.1778 | −1.8919 |
| 97.0 | −18.8333 | 16.7457 | −0.7280 | −0.4023 | −3.4263 | −1.8919 |

TABLE 7

$a = -10;\ p_1 = 52.2;\ q = 166$

| % CHORD | $\omega$(DEG.) | $\alpha$(DEG.) | CAL. $P$ | EXP. $P$ | CAL. $Q$ | EXP. $Q$ |
|---|---|---|---|---|---|---|
| 0.0 | 21.5388 | 52.4299 | 2.1615 | .... | 10.1729 | .... |
| 4.0 | 20.6000 | 50.7048 | 1.9810 | 1.9349 | 9.3234 | 9.1218 |
| 12.5 | 18.6278 | 47.5311 | 1.7158 | 1.6284 | 8.0752 | 7.6767 |
| 21.5 | 16.5445 | 44.5904 | 1.4415 | 1.3793 | 6.7843 | 6.5026 |
| 31.0 | 14.3612 | 41.8147 | 1.1880 | 1.1494 | 5.5912 | 5.4189 |
| 40.5 | 12.1178 | 39.2137 | 0.9502 | 0.9578 | 4.4720 | 4.5157 |
| 50.5 | 9.8333 | 36.7683 | 0.7311 | 0.7950 | 3.4408 | 3.7480 |
| 59.0 | 7.9388 | 34.8769 | 0.5650 | 0.5747 | 2.6591 | 2.7094 |
| 68.5 | 5.7555 | 32.8250 | 0.3892 | 0.4693 | 1.8317 | 2.2127 |
| 78.5 | 3.4055 | 30.7617 | 0.2182 | 0.3257 | 1.0269 | 1.5353 |
| 87.5 | 1.3278 | 29.0453 | 0.0811 | 0.1916 | 0.3817 | 0.9031 |
| 97.0 | −0.8333 | 27.6302 | −0.0281 | 0.0479 | −0.1322 | 0.2258 |

TABLE 8

$a = 14;\ p_1 = 52.2;\ q = 166$

| % CHORD | $\omega$(DEG.) | $\alpha$(DEG.) | CAL. $P$ | EXP. $P$ | CAL. $Q$ | EXP. $Q$ |
|---|---|---|---|---|---|---|
| 0.0 | − 2.4612 | 26.1665 | −0.1373 | .... | 0.6462 | .... |
| 4.0 | − 3.4000 | 25.5038 | −0.1853 | −0.1149 | −0.8721 | −0.5419 |
| 12.5 | − 5.8722 | 24.1589 | −0.2805 | −0.2586 | −1.3201 | −1.2192 |
| 21.5 | − 7.4555 | 22.8192 | −0.3713 | −0.3448 | −1.7475 | −1.6257 |
| 31.0 | − 9.6388 | 21.4984 | −0.4560 | −0.4406 | −2.1461 | −2.0772 |
| 40.5 | −11.8222 | 20.2539 | −0.5327 | −0.5172 | −2.5071 | −2.4385 |
| 50.5 | −14.1167 | 19.0280 | −0.6049 | −0.5747 | −2.8469 | −2.7094 |
| 59.0 | −16.0612 | 18.0449 | −0.6596 | −0.6034 | −3.1043 | −2.8449 |
| 68.5 | −18.2445 | 17.0079 | −0.7148 | −0.5172 | −3.3641 | −2.4385 |
| 78.5 | −20.5945 | 15.9651 | −0.7170 | −0.5172 | −3.3745 | −2.4385 |
| 87.5 | −22.6722 | 15.0954 | −0.8086 | −0.5172 | −3.8056 | −2.4385 |
| 97.0 | −24.8333 | 14.2418 | −0.8474 | −0.4980 | −3.9882 | −2.3482 |

TABLE 9

$a = 18;\ p_1 = 52.2;\ q = 166$

| % CHORD | $\omega$(DEG.) | $\alpha$(DEG.) | CAL. $P$ | EXP. $P$ | CAL. $Q$ | EXP. $Q$ |
|---|---|---|---|---|---|---|
| 0.0 | − 6.4612 | 23.4508 | −0.3289 | .... | −1.5479 | .... |
| 4.0 | − 7.4000 | 22.8533 | −0.3691 | −0.2490 | −1.7371 | −1.1741 |
| 12.5 | − 9.3722 | 21.6557 | −0.4462 | −0.3927 | −2.1000 | −1.8514 |
| 21.5 | −11.4555 | 20.4582 | −0.5202 | −0.4789 | −2.4483 | −2.2579 |
| 31.0 | −13.6388 | 19.2816 | −0.5901 | −0.5364 | −2.7772 | −2.5288 |
| 40.5 | −15.8222 | 18.1657 | −0.6529 | −0.6135 | −3.0728 | −2.8901 |
| 50.5 | −18.1167 | 17.0684 | −0.7116 | −0.6326 | −3.3491 | −2.9804 |
| 59.0 | −20.0612 | 16.2000 | −0.7554 | −0.5751 | −3.5552 | −2.7094 |
| 68.5 | −22.2445 | 15.2744 | −0.8001 | −0.5751 | −3.7656 | −2.7094 |
| 78.5 | −24.5945 | 14.3355 | −0.8432 | −0.5560 | −3.9684 | −2.6191 |
| 87.5 | −26.6722 | 13.5535 | −0.8770 | −0.5560 | −4.1275 | −2.6191 |
| 97.0 | −28.8333 | 12.7808 | −0.9087 | −0.5368 | −4.2767 | −2.5288 |

TABLE 10

$a = 20;\ p_1 = 52.2;\ q = 166$

| % CHORD | $\omega$(DEG.) | $\alpha$(DEG.) | CAL. $P$ | EXP. $P$ | CAL. $Q$ | EXP. $Q$ |
|---|---|---|---|---|---|---|
| 0.0 | − 8.4612 | 22.2013 | −0.4116 | .... | −1.9372 | .... |
| 4.0 | − 9.4000 | 21.6393 | −0.4472 | −0.3834 | −1.8044 | −1.8063 |
| 12.5 | −11.3722 | 20.5047 | −0.5174 | −0.4793 | −2.4351 | −2.2579 |
| 21.5 | −13.4555 | 19.3783 | −0.5844 | −0.5751 | −2.7504 | −2.7094 |
| 31.0 | −15.6388 | 18.2585 | −0.6477 | −0.6422 | −3.0483 | −3.0255 |
| 40.5 | −17.8222 | 17.3068 | −0.6974 | −0.6901 | −3.2822 | −3.2513 |
| 50.5 | −20.1167 | 16.1547 | −0.7577 | −0.6422 | −3.5660 | −3.0255 |
| 59.0 | −22.0612 | 15.3511 | −0.7965 | −0.5943 | −3.7486 | −2.7997 |
| 68.5 | −24.2445 | 14.4727 | −0.8370 | −0.5943 | −3.9393 | −2.7997 |
| 78.5 | −26.5945 | 13.5819 | −0.8758 | −0.5943 | −4.1219 | −2.7997 |
| 87.5 | −28.6722 | 12.8355 | −0.9066 | −0.5943 | −4.2668 | −2.7997 |
| 97.0 | −30.8333 | 12.1083 | −0.9354 | −0.5943 | −4.4024 | −2.7997 |

TABLE 11

$a = 26;\ p_1 = 52.5;\ q = 166$

| % CHORD | $\omega$(DEG.) | $\alpha$(DEG.) | CAL. $P$ | EXP. $P$ | CAL. $Q$ | EXP. $Q$ |
|---|---|---|---|---|---|---|
| 0.0 | −14.4612 | 18.8539 | −0.6146 | .... | −2.8926 | .... |
| 4.0 | −15.4000 | 18.3792 | −0.6410 | −0.5048 | −3.0168 | −2.3933 |
| 12.5 | −17.3722 | 17.4211 | −0.6928 | −0.5619 | −3.2606 | −2.6643 |
| 21.5 | −19.4555 | 16.4692 | −0.7419 | −0.6190 | −3.4917 | −2.9352 |
| 31.0 | −21.6388 | 15.5280 | −0.7880 | −0.6667 | −3.6616 | −3.1610 |
| 40.5 | −23.8222 | 14.6384 | −0.8295 | −0.7048 | −3.9040 | −3.3416 |
| 50.5 | −26.1167 | 13.7565 | −0.8685 | −0.6571 | −4.0875 | −3.1158 |
| 59.0 | −28.0612 | 13.0461 | −0.8982 | −0.6381 | −4.2273 | −3.0255 |
| 68.5 | −30.2443 | 12.3018 | −0.9275 | −0.6381 | −4.3652 | −3.0255 |
| 78.5 | −32.5945 | 11.5416 | −0.9559 | −0.6381 | −4.4988 | −3.0255 |
| 87.5 | −34.6722 | 10.8081 | −0.9789 | −0.6381 | −4.6071 | −3.0255 |
| 97.0 | −36.8333 | 10.2716 | −0.9998 | −0.6381 | −4.7055 | −3.0255 |

the top of each table, as well as the value of the *dynamic pressure* $q$ defined by $q = \frac{1}{2}\rho_1 W^2$ where $\rho_1$ is the density and $W$ the magnitude of the velocity of the flow in front of the shock. In addition to the calculated values of $P$(Cal. $P$), we have given the experimental values of this quantity (Exp. $P$) which are obtained directly from the tables of experimental values given by Ferri.[3] We have also given in these tables the calculated and experimental values of the quantity $Q$ (Cal. $Q$ and Exp. $Q$) defined by

$$Q = 15\left(\frac{\theta - p_1}{q}\right).$$

The experimental values of $Q$ are obtained from the tables given by Ferri by use of this relation. It is easily seen that the corresponding calculated values of $Q$ are given by the formula

$$\text{Cal. } Q = \frac{30(\text{Cal. } P)}{\gamma M^2}.$$

The graphs of Exp. $Q$ and Cal. $Q$ are constructed from the data in tables 2 to 11 and appear in figures 2 to 11. Figure 2 is obtained from table 2, figure 3 from table 3, etc. Full lines with experimental values indicated by small circles are used for the graphs of Exp. $Q$; dotted lines are used for the graphs of Cal. $Q$. In these graphs positive values of $Q$, which occur at points of the profile where the pressure is greater than the pressure in front of the shock, are measured upward from the heavy line and negative values of $Q$, where the pressure on the profile is less than the pressure before the shock, are measured downward from this line. The length of the side of a small square in the vertical direction, along which $Q$ is measured, represents $^1/_2$ unit; the length of the side of each large square (side lengths of four small squares) in the horizontal direction, along which per cent chord values are measured, represents 10 units. The left end of

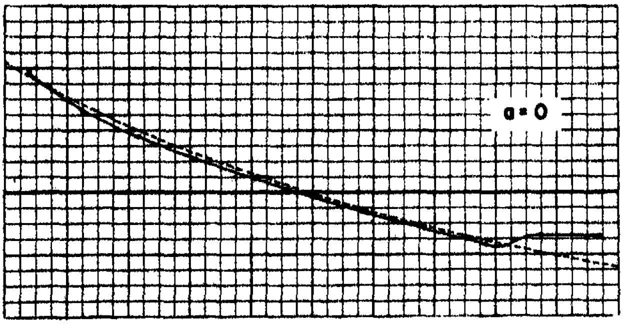


FIGURE 2

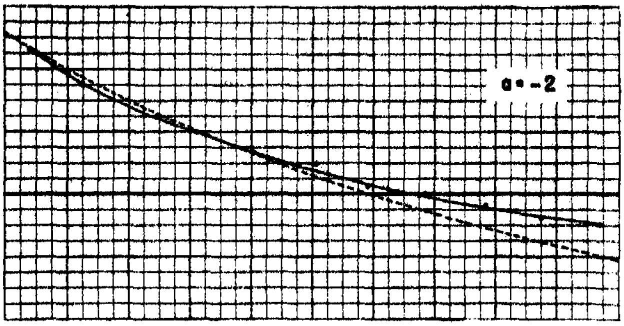


FIGURE 3

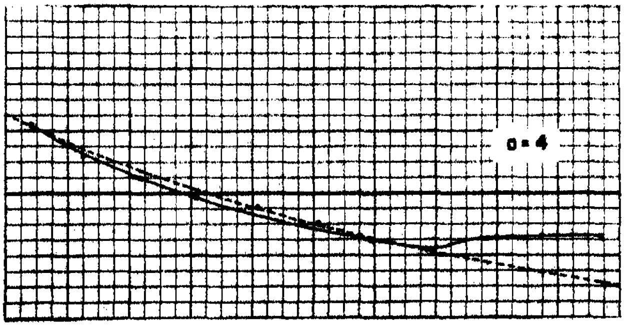


FIGURE 4

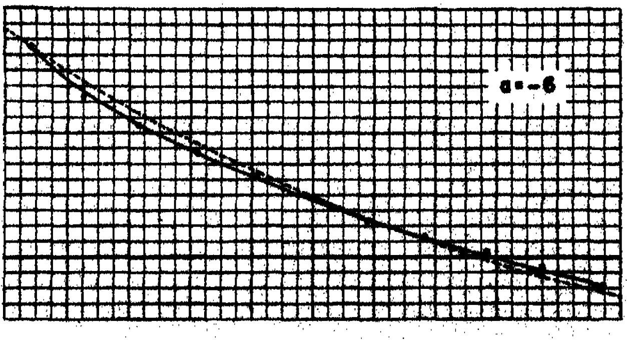


FIGURE 5

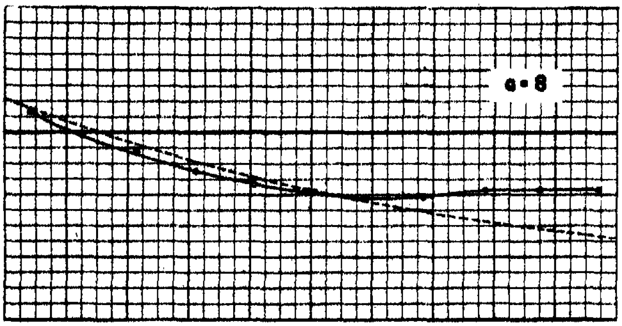


FIGURE 6

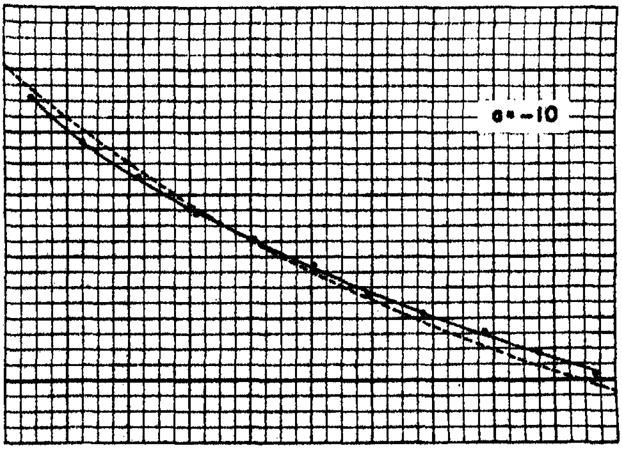


FIGURE 7

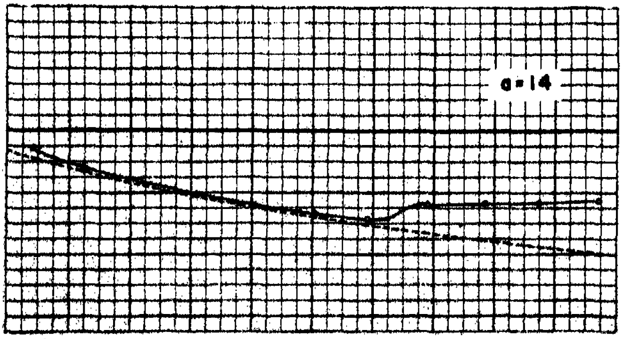


FIGURE 8

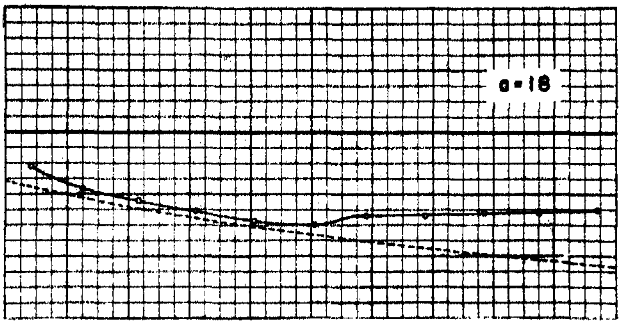


FIGURE 9

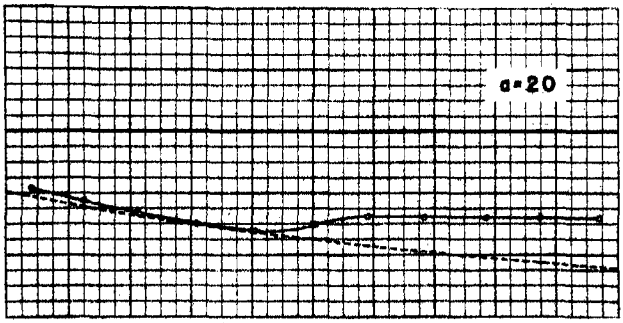


FIGURE 10

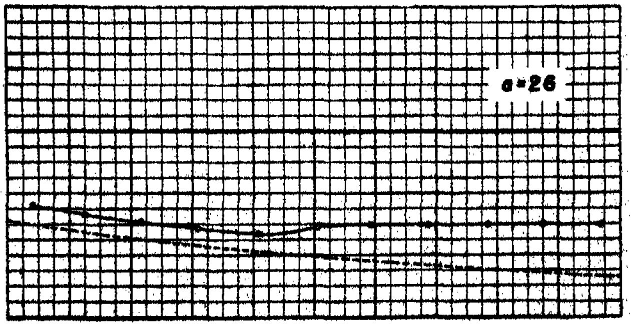


FIGURE 11

the heavy line corresponds to the 0% chord and the right end of this line to the 100% chord. The $Q$'s rather than the $P$'s or the direct-pressure ratios $\theta/p_1$ are used to permit a more direct comparison between our graphs and those given by Ferri, who has employed the quantity $Q$ for the graphical representation of his results.

*3. Conclusions.*—The graphs of the Cal. $Q$ show excellent agreement with the graphs, shown in the paper by Ferri, of the calculated values of this quantity obtained by the method of flow around a corner.[4] *Hence it appears that the method of pressure calculation based on the theory of flow around a corner can be replaced, for practical purposes, by the simple formula (1).*

At certain points of the profile where the pressure is less than the pressure $p_1$ and the flow is expanding, a shock wave originates which causes a sudden increase in the pressure as shown on the graphs. Since this method of pressure determination makes no provision for the occurrence of such shocks, there results a considerable deviation between the calculated and experimental values of $Q$. Moreover, at certain points of the profile preceding these shocks differences between the calculated and experimental values of $Q$ occur which appear too large to be attributed to experimental error. These differences can be accounted for in part by the presence of viscosity, attended by separation of flow from the boundary, which has been neglected in the theory, and in part by the obvious fact that the first approximation formula (1) cannot be expected to yield results of complete accuracy.

A determination of pressure distribution on curved profiles by the second approximation formula is now indicated. This involves the solution of an ordinary differential equation of the first order for the determination of the pressure function $\theta(\omega)$. A high-speed calculating machine is deemed advisable for the solution of this equation. If arrangements can be made for the use of such a machine the results on pressure distribution obtained will be presented in a later communication.

[1] "The Determination of Pressure on Curved Bodies Behind Shocks," to appear in *Communications on Applied Mathematics*, New York Univ. Inst. for Math. and Mech.

[2] Prepared under Navy Contract N6onr-180, Task Order No. 5, with Indiana University.

[3] Ferri, A., *Alcuni Risultati sperimentali riguardanti profili alari provati alla galleria ultrasonora di Guidonia*, Atti di Guidonia, No. 17, 1939, pp. 337–372. Translated as N.A.C.A., Technical Memorandum No. 946.

[4] See Courant, R., and Friedrichs, I. O., *Supersonic Flow and Shock Waves*, Interscience Publishers, New York, 1948, Chap. IVB, for a general discussion of the problem of flow around a corner. Specific procedures for pressure determination based on the method of flow around a corner have been given by Edmonson, Murnaghan and Snow, the Johns Hopkins University, Applied Physics Laboratory, Bumblebee Report, No. 26.

# *A GENIC SERIES CONTROLLING CHLOROPLAST PIGMENTS' PRODUCTION IN DIPLOID ZEA MAYS L.**

By Herbert L. Everett

Department of Genetics, Connecticut Agricultural Experiment Station, New Haven, Connecticut

Communicated by D. F. Jones, September 16, 1949

*Introduction.*—Inheritance of chlorophyll deficiencies controlled by single genes in *Zea mays* has been widely reported. Review articles by de Haan (1933)[1] and Eyster (1934)[2] describe many such cases of monofactorial control of plant pigment formation. Several studies have demonstrated the interaction of two or more genes in causing chloroplast pigment deficiencies. Lindstrom (1921)[3] described the interaction of two recessive factors termed *LLll* and *VVvv*, showing this type of inheritance:

| Gentotype | Phenotype |
|---|---|
| *LLvv* | Virescent—white |
| *llVV* and *LLVV* | Green—normal |
| *llvv* | Yellow |

Bach (1920),[4] Rasmusson (1920)[5] and Stroman (1924)[6] have also reported the existence of interacting genes which may be considered duplicate or supplementary in action. In more modern terms these would be termed suppressor gene mutants characterized by 15:1 ratios in segregating progenies. Demerec (1923, 1924, 1925, 1926)[7-10] carried out the most detailed studies of albino linkage and inheritance in maize that have been published to date. Demerec described duplicate and presumably triplicate genes for the formation of normal chloroplast pigments. Working with maize seed stocks which were segregating white and normal seedlings, Demerec demonstrated 15:1 ratio segregations in two stocks and later obtained two pedigrees which segregated in 63:1 ratios. The former condition indicates that two factors act as duplicate genes producing white seedlings in the double recessive only, while in the latter case, possibly three duplicate genes are present.

In an intriguing review of studies on the inheritance of chlorophyll, Demerec (1935)[11] suggested that qualitative differences as well as quantitative differences may occur in the various plant pigments. Such qualitative differences, it was suggested, might account for the variations in color which the maize breeder encounters as an everyday experience in observations of normal nurseries of corn inbreds (or families).

Recently, several studies concerning the inheritance of plastid pigments in *Zea mays* have attacked the problem of "physiological" or, more correctly, multigenic control of pigment formation. Elaborating upon the earlier investigations of Emerson (1929),[12] Sprague and Shive (1929)[13] and Jenkins (1929),[14] Sprague and Curtis (1933)[15] studied the chlorophyll content as an index of the productive capacity of selfed lines of corn and their hybrids. Sprague and Curtis concluded that "mean values for chlorophyll concentration and total chlorophyll of selfed lines are reasonably reliable indices of the total yield that will be produced by their hybrids." This work, while highly suggestive, has not been entirely confirmed by the more recent studies of Miller and Johnson (1938).[16] Their research utilized larger groups of material selected from the Minnesota Corn Improvement Program. After pigment extraction and spectrophotometric determinations, evidence was obtained that a highly significant positive correlation existed between the total chlorophyll and total carotinoids in the parental inbred lines and in their $F_1$ crosses. However, no significant correlation could be established between the total chlorophyll concentration in the leaf tissue and the grain yield of these hybrids. Further work by Miller and Johnson (1939, 1940)[17, 18] has shown that there is no correlation between the concentration of carotinoid pigments in seed endosperm of an inbred and that in the leaf tissue of the same inbred. Moreover, maternal inheritance was shown to be insignificant when reciprocal crosses of both high and low chlorophyll inbred lines were measured for chlorophyll content. Apparently, the male and female parents contributed equally to the $F_1$ crosses in respect to chlorophyll concentration. Hence, genotype in these studies determines phenotype without significant cytoplasmic or plastid influence.

*Experimental Materials.*—The problem of chlorophyll inher tance involves an understanding of the chloroplast pigments, their genic control and their correlation to plant production. Six *Zea mays* inbreds of varying phenotypic chlorophyll appearance were chosen: Dark (inbreds designated *C*106 and *T*1), Medium (*A*158 and *C.I.* 4-8) and Light (*W*23 and *Os*420). Chloroplast pigments were extracted and measured from seedlings of both the selected inbreds and $F_1$ generation seedlings of these inbreds crossed with a tester stock. This tester stock contributed both a pleiotropic gene and a suppressor gene mutation to the $F_1$ chlorophyll test crosses. Since an understanding of these genic types as well as the duplicate gene type is necessary for a clear comprehension of the results and discussion to be presented, the following definitions are given (after Knight, 1948)[19]:

*Pleiotropism:* The controlling of more than one character by a single gene. (This may also be termed Genuine Pleiotropism.)

*Suppressor Mutation:* A sudden change occurring at a particular chromosomal locus such that a gene is formed capable of nullifying the effect of a particular major gene.

*Duplicate Gene:* Identical genes but situated at different loci and showing no cumulative effect.

*Polymeric Genes:* Non-allelic genes governing the same character and having a similar effect on it.

Extraction of the plant pigments followed the analytical methods and utilized the constants for calculating total amounts of pigment presented by Griffith and Jeffrey (1944).[20] However, use of Skellysolve-B, a purified petroleum ether, as the solvent for all plant pigments extracted was found to expedite the entire procedure and thus reduce the time involved, and permit a larger number of samples to be run.

Since the absorption maxima obtained from Beckman spectrophotometric determinations of pigment extracts tended to be the same or very similar to those utilized by Griffith and Jeffrey, and since reproducibility of results was good, it was felt that quantitative calculations based on their formulae probably give a fair approximation of the pigment amounts present in the original seedlings.

Data from field plantings to be presented in conjunction with the plant pigment extractions and the segregation ratio counts (from seedlings grown under conditions of constant light and temperature) were based on Latin squares prepared by Fisher and Yates (1938).[21]

*Experimental Results.*—Table 1 presents the data from extracts of the six field corn inbreds while table 2 shows data from extracts of these six inbreds crossed by the tester stock. This tester was chosen because the line carries a pleiotropic gene that causes a kernel endopserm color change, wherein the germinated light kernel type yields a chlorophyll deficient seedling, while the dark kernels germinate to give perfectly normal plants.

TABLE 1*

| INBRED | USUAL PHENOTYPE | CHLOR. *a*, MG./G. | CHLOR. *b*, MG./G. | TOTAL CHLOR., MG./G. | % *a* CHLOR. | TOTAL CAROTENE, MG./G. | XAN., K./G. |
|---|---|---|---|---|---|---|---|
| *Os*420 | Light | 0.764 | 0.556 | 1.320 | 57.9 | 0.068 | 0.032 |
| *W*23 | Light | 0.613 | 0.138 | 0.751 | 81.6 | 0.040 | 0.025 |
| *A*158 | Medium | 0.699 | 0.102 | 0.801 | 87.3 | 0.057 | 0.026 |
| *C.I.* 4–8 | Medium | 0.688 | 0.262 | 0.950 | 72.4 | 0.047 | 0.023 |
| *T*1 | Dark | 0.768 | 0.303 | 1.071 | 71.7 | 0.070 | 0.027 |
| *C*106 | Dark | 0.762 | 0.456 | 1.218 | 62.6 | 0.055 | 0.030 |

* Average of duplicate samples—based on fresh weight.

A second gene which tends to overcome the chlorophyll suppression action of the pleiotropic gene is also present in the tester stock. These genes act independently of each other. Hence, the suppressor gene mutant segregates within the progeny of the light kernels and because it is incompletely dominant, three chlorophyll color classes are detectable in these plants. The classes are designated as albino (white), light-green (yellow-green) and pseudo-normal (practically a normal green). The suppressor gene does not affect the endosperm color variation brought about through the action of the pleiotropic gene nor does it overcome the lethal effect of the pleiotropic gene for after a few weeks the pseudo-normal seedlings take on a watery appearance and quickly die off.

The $F_2$ segregating progenies of the original inbred color lines crossed by the tester stock and subsequently selfed, showed that the light and medium inbreds (*W*23, *Os*420, *A*158, *C.I.* 4–8) contained normal alleles of the pleiotropic gene and recessive alleles of the suppressor gene mutant in the homozygous state. However, the dark inbreds (*T*1 and *C*106) were shown to have both the pleiotropic gene and the suppressor gene mutant in the homozygous state. In addition to the two chloroplast pigment controlling genes, it was seen that the dark lines contained a third completely dominant gene for normal chlorophyll production which incidentally tended to overcome the lethal effect of the pleiotropic gene.

TABLE 2*

| TEST CROSS | PHENOTYPES CROSSED | CHLOR. *a*, MG./G. | CHLOR. *b*, MG./G. | TOTAL CHLOR., MG./G. | % *a* CHLOR. | TOTAL CAROTENE, MG./G. | XAN., K./G. |
|---|---|---|---|---|---|---|---|
| *Os*420 × tester | Light × Medium | 0.641 | 0.306 | 0.947 | 67.7 | 0.052 | 0.020 |
| *W*23 × tester | Light × Medium | 0.531 | 0.118 | 0.649 | 81.8 | 0.038 | 0.018 |
| *A*158 × tester | Medium × Medium | 0.650 | 0.228 | 0.878 | 74.0 | 0.046 | 0.024 |
| *C.I.* 4–8 × tester | Medium × Medium | 0.671 | 0.217 | 0.888 | 75.6 | 0.041 | 0.023 |
| *T*1 × tester | Dark × Medium | 0.650 | 0.258 | 0.908 | 71.6 | 0.050 | 0.018 |
| *C*106 × tester | Dark × Medium | 0.573 | 0.494 | 1.067 | 53.7 | 0.049 | 0.020 |

* Average of duplicate samples—based on fresh weight.

TABLE 3

SEEDLING SEGREGATION—$F_2$ GENERATION OF SELFS OF INBRED *W*23 CROSSED BY THE TESTER STOCK (ONLY LIGHT KERNELS PLANTED)

| EAR NO. | NO. SEED PLANTED | NO. SEED GERMINATED | ALBINO | | LIGHT-GREEN | | PSEUDO-NORMAL | | $X^2$ |
|---|---|---|---|---|---|---|---|---|---|
| | | | OBS. | EXP. | OBS. | EXP. | OBS. | EXP. | |
| 1 | 50 | 50 | 10 | 12.50 | 28 | 25.00 | 12 | 12.50 | 0.88 |
| 2 | 50 | 50 | 12 | 12.50 | 24 | 25.00 | 14 | 12.50 | 0.24 |
| 3 | 50 | 48 | 10 | 12.00 | 23 | 24.00 | 15 | 12.00 | 1.12 |
| 4 | 50 | 47 | 9 | 11.75 | 23 | 23.50 | 15 | 11.75 | 1.55 |
| 5 | 50 | 47 | 18 | 11.75 | 19 | 23.50 | 10 | 11.75 | 4.44 |
| Totals | 250 | 242 | 59 | (60.50) | 117 | (121.00) | 66 | (60.50) | |

$X^2$ based on individual ears (expected ratio 1:2:1) = 8.23, where *D.F.* = 10 shows *P.* = 50–95%.

$X^2$ based on totals of all ears = 0.67, where *D.F.* = 2 shows *P.* = 50–95%.

This gene may be considered to be semi-duplicate to the pleiotropic gene in that it completely masks the latter's effect upon plant pigment production while the endosperm color effect is unchanged. Tables 3 and 4 show two representative counts of seedling types—one from a light type ($W23$) and one from a dark type inbred ($T1$).

TABLE 4

SEEDLING SEGREGATION—$F_2$ GENERATION OF SELFS OF INBRED $T1$ CROSSED BY THE TESTER STOCK (ONLY LIGHT KERNELS PLANTED)

| EAR NO. | NO. SEED PLANTED | NO. SEED GERMINATED | ALBINO-LIGHT GREEN OBS. | EXP. | PSEUDO-NORMAL OBS. | EXP. | NORMAL OBS. | EXP. | $X^2$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 50 | 50 | 0 | ... | 10 | 12.50 | 40 | 37.50 | 0.67 |
| 2 | 50 | 48 | 0 | ... | 9 | 12.00 | 39 | 36.00 | 1.25 |
| 3 | 50 | 49 | 0 | ... | 10 | 12.25 | 39 | 36.75 | 0.55 |
| 4 | 50 | 50 | 0 | ... | 11 | 12.50 | 39 | 37.50 | 0.24 |
| 5 | 50 | 50 | 0 | ... | 11 | 12.50 | 39 | 37.50 | 0.24 |
| 6 | 50 | 50 | 0 | ... | 13 | 12.50 | 37 | 37.50 | 0.03 |
| Totals | 300 | 297 | 0 | ... | 64 | (74.25) | 233 | (222.75) | |

$X^2$ based on individual ears (expected ratio 3:1) = 2.98, where $D.F.$ = 6 shows $P.$ = 50–95%.

$X^2$ based on totals of all ears = 1.88, where $D.F.$ = 1 shows $P.$ = 10–20%.

In the summer of 1948 the six $F_1$ test crosses (chlorophyll inbreds X tester stock) were entered in a 6 × 6 Latin square field trial. Rows were planted thickly and later thinned to a uniform stand so that variation in the total number of plants per entry was not a factor in the yields obtained. Single cross yields as well as other pertinent data from this trial are presented in table 5.

TABLE 5

YIELD TRIAL—PLANTED MAY 20, 1948; HARVESTED OCT. 29, 1948

| ENTRY | TOTAL PLANTS | *YIELD, POUNDS AT 15.5% MOISTURE CONTENT | MATURITY, DAYS |
|---|---|---|---|
| $T1$ × tester | 174 | 75.7 | 77 |
| $C106$ × tester | 166 | 67.5 | 77 |
| $Os420$ × tester | 169 | 63.5 | 76 |
| $A158$ × tester | 171 | 62.1 | 73 |
| $W23$ × tester | 167 | 61.8 | 75 |
| $C.I.$ 4-8 × tester | 171 | 56.2 | 79 |

* *Just significant difference* = 11.1 pounds at 0.05$P$.

Several significant correlations may be noted in the pigment extraction data and grain yield figures and these significant correlations may be seen to be dependent upon the pigment controlling genomes of the inbreds involved. These correlations may be summed up as follows:

| | VALUES COMPARED | DEGREES OF FREEDOM | "r" VALUES | SIGNIFICANCE |
|---|---|---|---|---|
| 1. | Total carotene (inbreds): total carotene (single crosses) | 4 | +0.934 | High positive correlation |
| 2. | Total chlorophyll (inbreds and single crosses): total carotene (inbreds and single crosses) | 10 | +0.733 | High positive correlation |
| 3. | Ratio of chlorophyll *a*/total carotene (inbreds and single crosses): yield in pounds of grain (single crosses only) | 10 | −0.639 | Significant negative correlation |

These correlations are in agreement with previous work but show in addition that the ratios between the chloroplast pigments are of significant importance in determining plant productivity (on a grain basis in this case).

*Discussion.*—The way in which a chloroplast is synthesized would seem to have a direct influence on the way in which it functions in the process of photosynthesis. This being true, the study of plants in which mutations have occurred in chloroplast synthesis is also a study of plants in which the photosynthetic capacities have been altered to a greater or lesser degree. The data presented demonstrate that even in inbred lines of diploid *Zea mays* the genes controlling the production of the chloroplast pigments are quite variable. Undoubtedly the discovery and isolation into separate progeny as shown in the segregating seedling types (tables 3 and 4) presage the finding of a series of genes controlling the formation of the chloroplast skeleton and, subsequently, the chloroplast pigments. It is suggested that such a series be designated Cl$_1$cl$_1$, Cl$_2$cl$_2$, Cl$_3$cl$_3$, . . . Cl$_x$cl$_x$. Thus, the three genes described may be symbolically represented as follows:

| TYPE OF GENE | SUGGESTED DESIGNATION |
|---|---|
| Pleiotropic | Cl$_1$cl$_1$ |
| Suppressor gene mutant | Cl$_2$cl$_2$ |
| Semi-duplicate | Cl$_3$cl$_3$ |

Utilizing the new series of designations, it is seen that originally the light and medium inbreds, *Os*420, *W*23, *A*158 and *C.I.* 4–8, had the genotype Cl$_1$Cl$_1$, cl$_2$cl$_2$, cl$_3$cl$_3$ while the dark inbreds, *T*1 and *C*106, had the genotype Cl$_1$Cl$_1$, Cl$_2$Cl$_2$, Cl$_3$Cl$_3$ and the tester stock genotype was Cl$_1$cl$_1$, Cl$_2$Cl$_2$, cl$_3$cl$_3$. These genotypes show that the dark lines vary from the light and medium inbreds in two of the three genes tested. This variation undoubtedly accounts for the fact that while plants of the light and medium lines are often transitional in color, the dark inbreds are always distinguished from the medium and light lines with ease. It is probable that unidentified modifying genes cause the color variation between the light and medium lines.

From this evidence, it seems probable that the variation in genes controlling the relative amounts of the chloroplast pigments accounts for the *distinct variation in inbred color and, through alteration of photosynthetic capacities, for the significant differences in grain production in the single crosses with the tester stock. Work is now in progress to determine* chromosome loci of the three genes described and to discover and isolate other genes of the pigment control series. These data as well as a more complete publication concerning the genetic, economic and evolutionary importance of this genic series will be prepared in the near future.

* An abstract of a dissertation presented for the degree of Doctor of Philosophy in Yale University.

[1] de Haan, H., *Bibliogr. Genet.*, **10**, 357–412 (1933).

[2] Eyster, W. H., *Ibid.*, **11**, 187–392 (1934).

[3] Lindstrom, E. W., *Genetics*, **6**, 91–110 (1921).

[4] Bach, S., *Zeitschr. Pfl. Zücht.*, **7**, 238–241 (1920).

[5] Rasmusson, J., *Hereditas*, **1**, 128–134 (1920).

[6] Stroman, G. N., *Genetics*, **9**, 91–117, 493–512 (1924).

[7] Demerec, M., *Bot. Gaz.*, **84**, 139–155 (1923).

[8] Demerec, M., *Abs. Anat. Rec.*, **29**, 133 (1924).

[9] Demerec, M., *Genetics*, **10**, 318–344 (1925).

[10] Demerec, M., *Am. Nat.*, **60**, 172–176 (1926).

[11] Demerec, M., *Cold Spring Harbor Symposia Quant. Biol.*, **3**, 80–86 (1935).

[12] Emerson, R., PROC. NATL. ACAD. SCI., **15**, 281–284 (1929).

[13] Sprague, H. B., and Shive, J. W., *Plant Physiol.*, **4**, 165–192 (1933).

[14] Jenkins, M. T., *J. Agr. Res.*, **39**, 677–721 (1929).

[15] Sprague, H. B., and Curtis, N., *J. Am. Soc. Agron.*, **25**, 709–724 (1933).

[16] Miller, E. S., and Johnson, I. J., *Ibid.*, **30**, 941–946 (1938).

[17] Johnson, I. J., and Miller, E. S., *Cereal Chem.*, **16**, 88–92 (1939).

[18] Miller, E. S., and Johnson, I. J., *Proc. Soc. Exp. Biol. and Med.*, **44**, 26–28 (1940).

[19] Knight, R. L., *Dictionary of Genetics*, Chronica Botanica Co., Waltham, Mass. 1948.

[20] Griffith, R. B., and Jeffrey, R. N., *Ind. Eng. Chem., Anal. Ed.*, **16**, 438–440 (1944).

[21] Fisher, R. A., and Yates, F., "*Statistical Tables for Biological, Agricultural and Medical Research*," Oliver and Boyd, London, 1938.

---

# *REPRODUCTION AND DISTRIBUTION OF THE CYTOPLASMIC FACTOR FOR MALE STERILITY IN MAIZE**

BY WARREN H. GABELMAN

DEPARTMENT OF GENETICS, CONNECTICUT AGRICULTURAL EXPERIMENT STATION,
NEW HAVEN, CONNECTICUT

Communicated by D. F. Jones, September 8, 1949

Recent work of Preer,[1, 2] Sonneborn[3, 4] and L'Héritier,[5, 6] has emphasized and helped clarify the rôle of the cytoplasm in the determination and inheritance of characters. The present paper deals with the inheritance of a cytoplasmic factor in maize in such a way as to determine particle number, rate of reproduction of the particle and the distribution of the particles within the plant. This is the same cytoplasmic factor previously studied by Rhoades.[7] Similar cytoplasmic factors causing male sterility have been found in onions by Jones and Clarke[8] and in sugar beets by Owen.[9] In both of these latter cases, an interaction between the cytoplasmic factor and chromogenes was reported.

Rhoades[7] showed that microsporogenesis was normal and that the abortion occurred during the development of the microspore into a mature male gametophyte. If the plant was partially fertile, certain of these microspores would continue development normally. When this partially fertile pollen was used in cross-pollinations there was no evidence of a transfer of the male sterile factor. This fact suggested the possibility of the presence of one or more cytoplasmic factors in a microspore destined to abort and the complete absence of the factor in those microspores which would develop into normal pollen grains.

Partially fertile lines were selected for a detailed study of variability in pollen production among florets of a given plant. Such a study would rule out the possibilities of genic and physiological differences which might normally arise in a comparison between plants. The use of partially fertile plants makes possible the detection of small differences not measurable in completely sterile lines.

The amount of viable pollen was determined by smearing the pollen from the anthers of a single floret in a drop of aqueous iodine-potassium iodide solution. Those pollen grains which were plump and full of starch were considered viable. If there were any shrunken, starch-deficient or otherwise aborted areas in the pollen grain, it was considered to be non-viable. Cross-pollinations using the partially fertile types as pollen parents were successful in direct proportion to the amount of viable pollen shed per plant as determined by this method.

Five tassels from partially fertile plants were chosen at random. The per cent of viable pollen was determined by observing 150–200 pollen grains per floret, using a Breed micrometer to limit the amount of the field counted. Counts were made at a magnification of 150 ×. Five such counts at random usually included 150–200 pollen grains. When such data from any individual tassel were subjected to a $2 \times N$ chi square test, the variation among the florets exceeded that expected by chance ($P = <0.001$). Frequency distributions were made of the percentages of good pollen produced in these florets. Such frequency distributions were bimodal or polymodal with only one approaching unimodality. The modes of these distributions were at 36 and 60 per cent viable pollen. Figure 1 presents these data combined to show these peaks more clearly.

Each of the five plants used in figure 1 when studied by itself was free of genic and physiological differences so that such polymodal differences must have been due to the cytoplasmic factor itself. Since the modes from these five plants were at the same points it would indicate that the same system was involved in each case.

An explanation for such a distribution may be formulated in the following manner. Assume the presence of *k* discrete cytoplasmic particles in the microsporocytes which give rise to aborted pollen grains. Since functional pollen produced by a partially fertile plant does not transfer the cytoplasmic factor it can be assumed to be free of the male sterile particle. It follows that the aborted pollen grains contain one or more of these particles. The particle number must remain constant through successive cell divisions. If, at mitosis, the particles were distributed to daughter cells by chance alone, cells should arise devoid of the particle. The descendents of such a cell should all be devoid of particles and a pollen fertile chimera should result. No such chimeras have been found in any of the plants tested. From inspection of the variability in figure 1, it is evident that at some late cell division, such as during meiosis, the control which previously had insured equal division must be impaired. If this distribution at meiosis were random, a Poisson series would be expected.

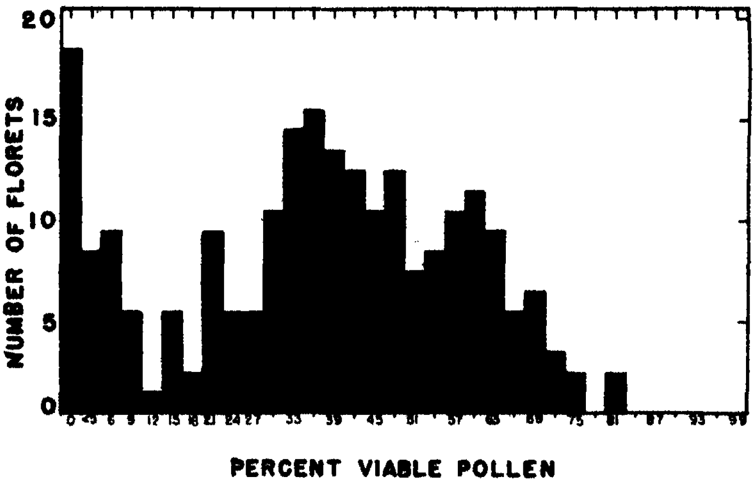


FIGURE 1

Frequency distribution of number of florets plotted against percentage of viable pollen, grouped in intervals of 3 per cent, accumulated from 5 plants.

The Poisson series is the basic discontinuous random distribution. It may occur when a variable is not able to take all possible values, but is confined to a particular series of values, such as whole numbers. If a variate can take the values 0, 1, 2, . . ., $X$, . . ., and the relative frequencies with which the values occur are given by the series

$$e^{-m}\left(1, m, \frac{m^2}{2!}, \ldots, \frac{m^x}{x!}, \ldots\right)$$

then the number is distributed in the Poisson series.[10] The mean is the only unknown parameter of the Poisson series. If $m$ is the mean number of particles per pollen grain, the proportion of samples containing none is $e^{-m}$. The pollen grains, free of the particle, would be viable. If each peak (Fig. 1) represented the mean of a component Poisson series, the average number of particles per pollen grain would be $^1/_2$ and 1 for 60 and 36 per cent viable pollen, respectively.

If the number of particles per cell were constant throughout the plant and only two types of pollen could be identified, pollen free of the particles and pollen containing one or more particles, it should be possible to plot the percentage of viable pollen in successive samples as a unimodal frequency distribution. Figure 1 is polymodal although only two modes predominate.

Two possibilities may explain such a divergence from the Poisson. Occasionally particle division and cell division might not coincide during mitoses, leading to an increase or decrease in the number of particles per daughter cell. Alternately, the distribution of particles to daughter cells may not always be exact.

Either of these possibilities could lead to a heterogeneous population of particles within the cells of a single plant. Random samples drawn from such a population would take the form of a negative binomial distribution.[11] In experimental sampling the negative binomial arises from a single extension of the conditions underlying the Poisson series. The Poisson series arises when equal samples are taken from perfectly homogeneous material. It is completely determined by the average number of particles, $m$ per sample. If unequal samples were taken, or, if the material was not perfectly homogeneous, the value of $m$ would vary from sample to sample.

If an atypical division happens only occasionally, a stage should be reached as meiosis is approached where a significant discrepancy in particle number and consequently in the percentage of viable pollen would be expected in some cases and not in others. This can be tested by comparing the percentages of good pollen in the anthers of the two florets which comprise the spikelet. Although still many cell divisions from meiosis, the anthers represent a much later common derivation than can be claimed for florets on different spikelets. In consequence, a $2 \times N$ chi square test should show agreement with the binomial in some cases and not in others. Comparison of the viable pollen in the anthers of the five spikelets studied showed probabilities of 0.78, 0.22, $<0.01$, $<0.001$ and $<0.001$. This shows that a binomial distribution is fit occasionally as meiosis is approached which is in agreement with the proposed hypothesis.

If the male sterile factor is distributed at random during microsporogenesis, it is of fundamental interest to see what type of distribution is had during megasporogenesis. Is this random distribution a function of meiosis?

This problem was tested by studying six florets (at random) from each of 78 plants from a three-way cross. These data are shown graphically in figure 2.

In making such counts the accuracy is very limited below 10 per cent and above 90 per cent. Actual counts were only made on those florets about 5 per cent or more viable. It should be noted that three definite peaks can be found at 15, 36 and 57 per cent. This is in agreement with mean particle numbers of 2, 1 and $^1/_2$, respectively. It should also be noted that the variability is greater in figure 2 than in figure 1, which indicates that the random distribution probably took place at megasporogenesis of the parent plant as well as microsporogenesis of the progeny and is a function of meiosis.

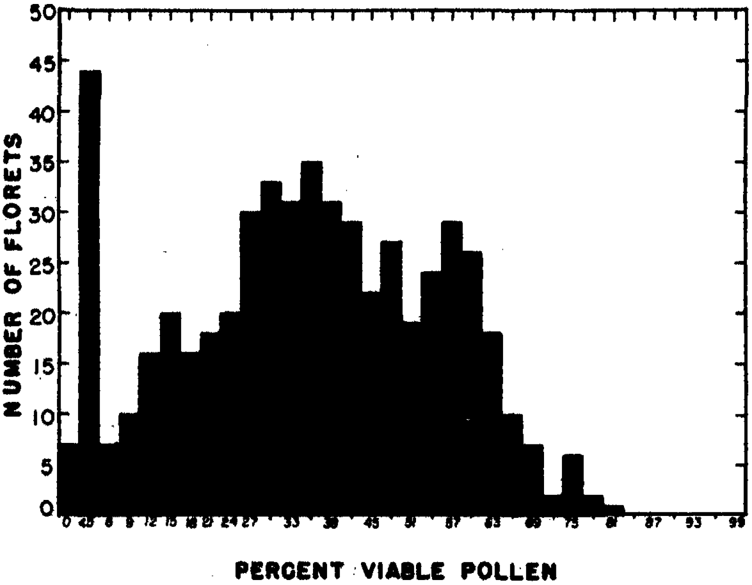


FIGURE 2

Frequency distribution of number of florets plotted against percentage of viable pollen, grouped in intervals of 3 per cent, accumulated after determining per cent of good pollen in 6 florets at random on each of 78 plants.

The female plant in the three-way cross used in this study produced 49 per cent viable pollen in 1947. This corresponds to 0.714 particle per pollen grain. The median of the florets (Fig. 2) was 0.741 particle. In a second study the female had 0.891 particle per pollen grain (1947) and the progeny 0.967 particle per pollen grain (1948). These data suggest that there was no appreciable increase or decrease in the number of particles per cell during the life cycle of these two families.

One of the basic premises of the Poisson series is that the variate must be confined to a particular series of values, such as whole numbers. If the number of particles per cell must be one or more (as shown previously by the absence of pollen fertile chimeras), a mean number of $^1/_2$ needs to be justified.

The assumption that the particle number is relatively constant at mitosis seems to be valid. It has also been assumed that at some stage, presumably at meiosis, the forces in control during mitosis are absent and distribution of the particle becomes random. Since the division of particles presumably paralleled that of the chromosomes during mitosis a similar parallelism during meiosis would suppress particle division at the same time and lead to the random distribution of undivided particles. If the cytoplasmic particles were to divide only once during meiosis as do the chromosomes, a mean number of 1 per somatic cell will be reduced to $^1/_2$ per microspore. Similarly, peaks arose at $^1/_2$ and 1 particle per microspore. Never was there any variance from this distribution. Therefore, the numerical difference between these peaks ($^1/_2$) should be the equivalent of a whole particle per somatic cell. Similarly modes were not found between 0 and $^1/_2$ particle per cell.

*Discussion.*—The significance of plasmagenes has been the subject of much speculation. Sonneborn[12] has recently reviewed the numerous views presented by many workers. From the evidence presented here it is possible to characterize partially the cytoplasmic factor causing male sterility in corn.

The male sterile factor is particulate. If it were not a whole unit or particle its distribution would not follow a negative binomial made up by a series of Poisson distributions.

The reproduction and distribution of this particle are quite similar to the reproduction and distribution of chromosomes. The similarity in reproduction holds both at meiosis and mitosis. The similarity of distribution holds quite well at mitosis but is completely random at meiosis. It must, therefore, have many characteristics in common with the chromosome. It might also be suggested that this is a virus. All attempts to transfer it artificially have failed. It is doubtful that a virus would be so dependent on chromosome division for its reproduction and distribution.

---

[*] This paper is a part of a dissertation presented for the degree of Doctor of Philosophy at Yale University.

[1] Preer, J. R., Proc. Nat. Acad. Sci., **32**, 247–253 (1946).

[2] Preer, J. R., *Genetics*, **33**, 349–404 (1948).

[3] Sonneborn, T. M., PROC. NATL. ACAD. SCI., **29**, 329–343 (1943).

[4] Sonneborn, T. M., *Adv. in Genetics*, **1**, 264–358 (1947).

[5] L'Héritier, P., and Teissier, G., *Compt. rend.*, **206**, 1193–1195 (1938).

[6] L'Héritier, P., *Heredity*, **2**, 325–348 (1948).

[7] Rhoades, M. M., *J. Gen.*, **27**, 71–93 (1933).

[8] Jones, H. A., and Clarke, A. E., *Proc. Am. Soc. Hort. Sci.*, **43**, 189–194 (1943).

[9] Owen, F. J., *J. Agr. Res.*, **71**, 423–440 (1945).

[10] Fisher, R. A., *Statistical Methods for Research Workers*, Oliver and Boyd, Edinburgh, 1948.

[11] Fisher, R. A., *Ann. Eugen.*, **2**, 182–187 (1941).

[12] Sonneborn, T. M., *Am. Sci.*, **37**, 33–59 (1949).

---

# *THE EFFECT OF OXYGEN ON THE FREQUENCY OF X-RAY INDUCED CHROMOSOMAL REARRANGEMENTS IN TRADESCANTIA MICROSPORES**

BY NORMAN H. GILES, JR., AND HERBERT PARKES RILEY†

BIOLOGY DIVISION, OAK RIDGE NATIONAL LABORATORY

Communicated by Karl Sax, August 31, 1949

The recent studies of Thoday and Read,[1] using root tips of *Vicia faba*, indicate that the availability of oxygen to cells is a very important factor in radiosensitivity. Under anaerobic conditions, sensitivity to x-rays is markedly reduced, as measured both by inhibition of growth and by frequency of anaphase figures showing chromosome aberrations. The results of Hayden and Smith[2] on x-radiation of barley seeds in a vacuum also suggest that the absence of oxygen results in reduced radiosensitivity. In view of these results it seemed of interest to test the effect of oxygen on the frequency of x-ray induced chromosomal rearrangements in Tradescantia, especially since there is already available a wealth of information on radiation effects in this organism as a result of the pioneer investigations of Sax[3] which, together with those of other investigators, have been summarized by Lea[4] and by Catcheside.[5]

*Materials and Methods.*—*Tradescantia paludosa* Anderson and Woodson, clone 5 of Sax, was used in all experiments. Observations were made at the first postmeiotic mitosis in the microspore by means of acetocarmine smear preparations at the four- to five-day interval following treatment, at which time only chromosome-break types are present. The principal aberration types analyzed at this period are interchanges (dicentrics and centric rings). In addition, interstitial deletions (intercalary deletions or isodiametric fragments) are numerous. Acentric rings and terminal deletions were also recorded but, because of their comparative rarity, were not used in later calculations.

The x-ray source, a Coolidge self-rectifying tube with tungsten target was operated at 250 kvp. and 15 ma. The inherent filtration was equivalent to 3 mm. aluminum. In addition, all exposures were made inside a lucite exposure chamber whose top was $^1/_4$-inch thick. The inflorescences were inserted in an upright position through holes in a very thin lucite platform in the center of the box in order to insure uniform exposures and to avoid back-scattering. Dosages were measured inside the exposure chamber at the position of the inflorescences with a 100-r Victoreen thimble ionization chamber recently calibrated by the U. S. Bureau of Standards. In all experiments, unless otherwise noted, the dosage rate was constant, 45 r/min.–different total dosages being administered by increasing the time of exposure.

In order to make exposures in different gases, inflorescences were first placed in a suction flask, evacuated with a water pump for approximately two and one-half minutes, and the appropriate gas then admitted into the flask. This procedure was repeated five times. The purpose of this operation was to remove the air enclosed by the sepals and petals of the buds. The inflorescences were then quickly placed in the lucite exposure chamber, which could be made airtight, the air evacuated and the appropriate gas admitted. This operation was performed twice. The lucite chamber was next placed in the x-ray machine and the inflorescences irradiated at room temperature. After exposure, the inflorescences were kept in the gas chamber for approximately ten minutes, following which they were removed, reëvacuated in the suction flask and air permitted to diffuse in. This procedure was repeated five times, as previously. Commercial cylinders of the various gases were utilized, the indicated purity in all cases being at least 99.5%.

In the first control experiment in which inflorescences were exposed to x-rays in air inside the lucite chamber, the same initial series of evacuations was carried out, except that air rather than some other gas was permitted to enter the flask following each evacuation. A repetition of this control experiment in air without prior evacuation of the buds gave essentially similar results indicating that the evacuation procedure itself did not modify the results. Observations of other control buds, exposed to atmospheres of nitrogen, oxygen and other gases in an identical manner but not x-rayed, indicated that exposure to the gas alone did not result in any detectable cytological abnormalities.

In general from 10 to 15 inflorescences were exposed at each treatment and as many slides as possible made at the four- and five-day interval following irradiation. The aberration frequency for each dosage was calculated from the total number of interchanges and interstitial deletions recorded in either 50 or 100 cells on 3 to 15 (average ca. 7) slides. Chi-square tests did not reveal any excess variation among slides from one treatment and standard errors were calculated by the method discussed by Catcheside, Lea and Thoday.[6]

*Results and Discussion.*—The results of comparative experiments in which inflorescences were irradiated in air, in nitrogen and in oxygen are presented in table 1 and shown graphically in figures 1 and 2. All the data in experiment 1 (exposures in air) were obtained by one investigator (N. H. G.). Combined data from the scorings of both investigators for interchanges (dicentrics and centric rings) are given in all succeeding experiments since the scorings of these aberration types by both investigators are quite comparable. However, there is considerable individual variation in the scoring of interstitial deletions because of the small size of some of these aberration types and the more extensive data of only one investigator (H. P. R.) have been used in all except the first experiment. In general, these values are lower than those obtained in comparable scorings by the other author (N. H. G.) by approximately 20%. Despite this, the relative differences from one experiment to another are consistent in both sets of data.

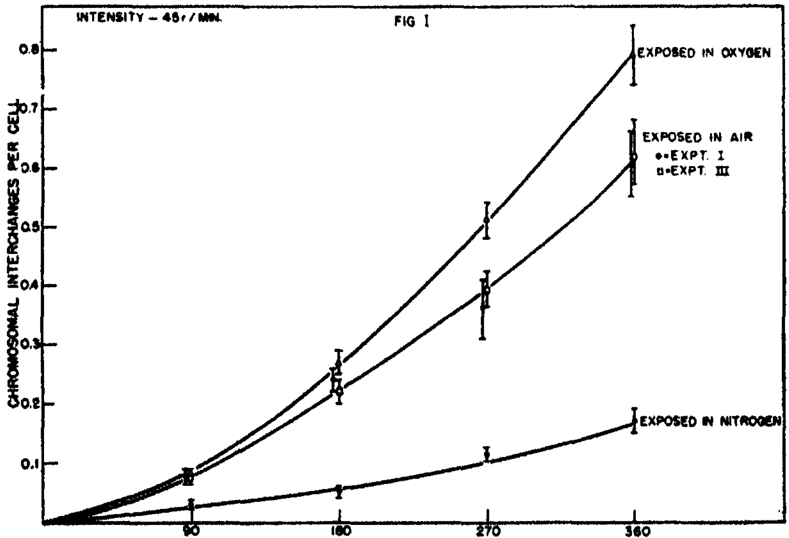


FIGURE 1

X-ray dosage curves for interchanges (dicentrics and centric rings) induced in Tradescantia microspores exposed in atmospheres of air, nitrogen and oxygen. Vertical bars indicate limits of standard errors.

It is clear from these data that the aberration frequency is strikingly reduced when exposures are made in nitrogen instead of in air. Further, there is a significant though not as marked a difference when oxygen is used in place of air, the result in this instance being an increased aberration yield. The magnitude of these differences is very considerable, the increase in both interchanges and deletions in exposures made in oxygen as compared with nitrogen being as much as fivefold at higher dosages. These results strongly suggest that the presence of oxygen has a marked influence in increasing radiosensitivity in this material as measured by chromosomal aberration frequency.

There is no reason to believe that the observed effects can be accounted for by a modification of the timing of the cell cycle as a result of exposure to the different gases such that variations in mitotic sensitivity are involved. Observations of the x-ray effects were not made until the fourth and fifth days following treatment, by which time any effect of a possible acceleration or retardation of the mitotic cycle should have little if any consequence. It has been shown by Sax[8] that there is very little change in x-ray sensitivity throughout the five- to six-day-long resting stage. Further, there is no evidence from the present data of significant differences in aberration frequencies in comparisons of four- and five-day material within individual experiments.

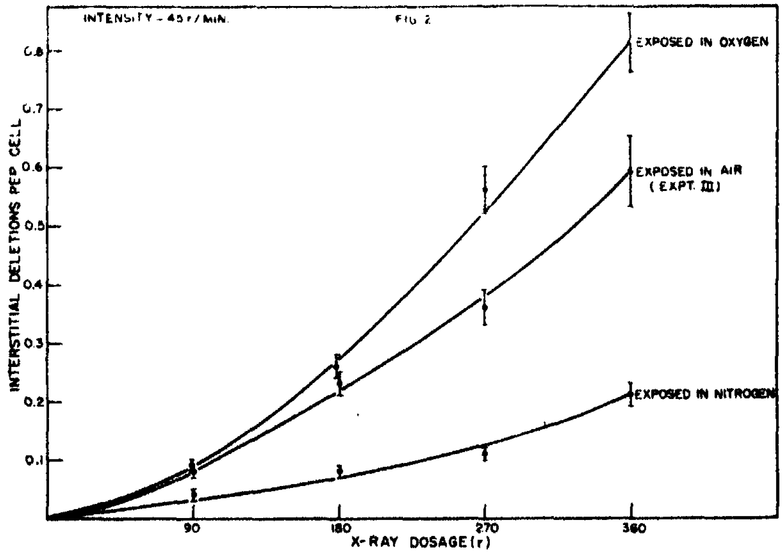


FIGURE 2

X-ray dosage curves for interstitial deletions induced in Tradescantia microspores exposed in atmospheres of air, nitrogen and oxygen. Vertical bars indicate limits of standard errors.

In order to eliminate the possibility that the presence of nitrogen rather than the absence of oxygen might be responsible for the observed decrease in radiosensitivity, further experiments were performed in which other gases were substituted for nitrogen. The results of such an experiment are presented in table 2. In this instance, comparative exposures were made at a single dose, 360 r at 45 r/min., in nitrogen, in helium and in argon. It will

TABLE 1

FREQUENCIES OF INTERCHANGES AND INTERSTITIAL DELETIONS INDUCED IN TRADESCANTIA MICROSPORES EXPOSED TO X-RAYS IN AIR, IN NITROGEN AND IN OXYGEN ALL DOSAGES ADMINISTERED AT A CONSTANT INTENSITY OF 45 R/MIN.
(For further discussion, see text)

| DOSE | CELLS SCORED | DICENTRICS | CENTRIC RINGS | TOTAL INTERCHANGES | INTERCHANGES PER CELL | CELLS SCORED | INTERSTITIAL DELETIONS | INTERSTITIAL DELETIONS PER CELL |
|---|---|---|---|---|---|---|---|---|
| | | | | EXPOSED IN AIR (EXPT. 1) | | | | |
| 90 | 500 | 29 | 11 | 40 | 0.08 ± 0.01 | 500 | 48 | 0.10 ± 0.01 |
| 180 | 276 | 45 | 20 | 65 | 0.24 ± 0.03 | 276 | 71 | 0.26 ± 0.03 |
| 270 | 150 | 44 | 10 | 54 | 0.36 ± 0.05 | 150 | 76 | 0.51 ± 0.06 |
| 360 | 200 | 92 | 29 | 121 | 0.61 ± 0.06 | 200 | 144 | 0.72 ± 0.06 |
| | | | | EXPOSED IN NITROGEN | | | | |
| 90 | 1150 | 23 | 9 | 32 | 0.03 ± 0.005 | 800 | 32 | 0.04 ± 0.01 |
| 180 | 400 | 15 | 4 | 19 | 0.05 ± 0.01 | 400 | 30 | 0.08 ± 0.01 |
| 270 | 1267 | 107 | 40 | 147 | 0.12 ± 0.01 | 800 | 90 | 0.11 ± 0.01 |
| 360 | 450 | 59 | 20 | 77 | 0.17 ± 0.02 | 450 | 93 | 0.21 ± 0.02 |
| | | | | EXPOSED IN AIR (EXPT. 3) | | | | |
| 90 | 1050 | 52 | 30 | 82 | 0.08 ± 0.01 | 800 | 65 | 0.08 ± 0.01 |
| 180 | 600 | 88 | 44 | 132 | 0.22 ± 0.02 | 600 | 139 | 0.23 ± 0.02 |
| 270 | 600 | 165 | 73 | 238 | 0.40 ± 0.03 | 400 | 143 | 0.36 ± 0.03 |
| 360 | 150 | 68 | 25 | 93 | 0.62 ± 0.06 | 150 | 88 | 0.59 ± 0.06 |
| | | | | EXPOSED IN OXYGEN | | | | |
| 90 | 800 | 50 | 15 | 65 | 0.08 ± 0.01 | 800 | 72 | 0.09 ± 0.01 |
| 180 | 900 | 165 | 77 | 242 | 0.27 ± 0.02 | 900 | 231 | 0.26 ± 0.02 |
| 270 | 500 | 175 | 78 | 253 | 0.51 ± 0.03 | 350 | 195 | 0.56 ± 0.04 |
| 360 | 300 | 157 | 80 | 237 | 0.79 ± 0.05 | 300 | 244 | 0.81 ± 0.05 |

TABLE 2

FREQUENCIES OF CHROMOSOMAL INTERCHANGES AND INTERSTITIAL DELETIONS INDUCED IN TRADESCANTIA MICROSPORES EXPOSED TO 360 R AT 45 R/MIN. IN ATMOSPHERES OF NITROGEN, ARGON AND HELIUM

| TREATMENT | CELLS SCORED | DICENTRICS | CENTRIC RINGS | TOTAL INTERCHANGES | INTERCHANGES PER CELL | CELLS SCORED | INTERSTITIAL DELETIONS | INTERSTITIAL DELETIONS PER CELL |
|---|---|---|---|---|---|---|---|---|
| 360 r in nitrogen | 500 | 96 | 36 | 132 | 0.26 ± 0.02 | 150 | 56 | 0.19 ± 0.05 |
| 360 r in argon | 500 | 130 | 50 | 180 | 0.36 ± 0.03 | 300 | 164 | 0.27 ± 0.04 |
| 360 r in helium | 300 | 38 | 14 | 52 | 0.17 ± 0.02 | 150 | 40 | 0.13 ± 0.04 |

be noted that helium is even more effective than nitrogen in this experiment in reducing the frequency of interchanges and deletions, the values obtained in this gas being as low as or lower than those found in the nitrogen exposure at 360 r in the first series of experiments. The value for interchanges in nitrogen is somewhat higher in this experiment than in the previous one, but is still markedly less than that obtained in air. The aberration frequencies in argon, although much lower than those in air, are higher than those obtained in nitrogen or helium. The reason for this difference is not apparent but is probably due to uncontrolled variables in the experimental techniques. It is clear from this experiment, however, that the reduced aberration frequencies as compared with air cannot be attributed to the presence of a particular gas, such as nitrogen, but must result from the absence of oxygen.

As yet no exposures in atmospheres of oxygen of known intermediate percentage composition, such as 5 or 10%, have been attempted. However, some experiments were performed in which the oxygen percentage was probably close to these intermediate values. These results were inadvertently obtained when an undiscovered air leak was present in the lucite exposure box resulting in a mixture of air and the gas being introduced. The interchange frequencies per cell at 360 r (45 r/min.) under these conditions were all intermediate between those obtained in exposures made in air and in the pure gases. These values were as follows: nitrogen, 0.43 ± 0.03; helium, 0.54 ± 0.03; argon, 0.52 ± 0.03 and 0.44 ± 0.03 (different experiments). The frequencies of interstitial deletions were also intermediate between those obtained in air and in the pure gases. The values (H. P. R. only) were: nitrogen, 0.53 ± 0.04; helium, 0.41 ± 0.03; argon, 0.42 ± 0.04 and 0.32 ± 0.03. It thus appears clear that the aberration frequency increases as the percentage of oxygen increases, but it is not yet possible to make a quantitative evaluation of this relation.

In attempting to interpret the mechanism by which oxygen increases the radiosensitivity of Tradescantia chromosomes, it is immediately necessary to decide whether this effect results from an increase in the initial frequency of x-ray-induced breaks or from an effect on the reunion of broken ends, such that recombination is favored over restitution. It is of course possible that both of these effects may be present. A preliminary analysis of the dosage curves for interchanges at a constant intensity of 45 r/min. utilizing Lea's[4] formulation for the recovery time, $\tau$,[7] suggests that the oxygen effect may be exerted principally by way of the recovery mechanism. However, it appears necessary to perform further exposures at higher and lower intensities in order to obtain more data on this point. Additional evidence is also being sought by means of another experimental approach. Previous investigations of a number of workers have indicated that in Tradescantia x-ray-induced chromosome breaks may remain open (i.e., available for participation in interchanges) for a considerable period of time. The average restitution time as calculated by Lea[4] is about four minutes. Thus it should be possible, if oxygen exerts its effect on the recovery mechanism so as to favor recombination, to increase the yield of rearrangements by transferring inflorescences irradiated at high intensity in nitrogen to oxygen as soon as possible after exposure. An initial attempt to perform such an experiment has not been conclusive. Buds were evacuated and nitrogen admitted in the usual fashion. The buds were then exposed to 300 r at 200 r/min. in the lucite box in air in order to facilitate removing half of the inflorescences to oxygen as rapidly as possible at the end of the irradiation. This transfer required approximately one minute. It was found that the frequency of interchanges was the same in both the control half (remaining in the lucite chamber) and in the experimental half (transferred to oxygen). The actual value obtained in the control group not transferred to oxygen was higher than expected as compared to comparable exposures in air at this intensity, however, and may indicate that the replacement of air by nitrogen in the buds alone is not sufficient, but that it is necessary to make the exposure in an atmosphere devoid of oxygen as well. This problem is being investigated further.

*Summary.*—Experiments have been performed which indicate that oxygen has a marked effect in increasing the radiosensitivity of Tradescantia chromosomes. Comparative exposures of inflorescences to x-rays were made in air, in nitrogen and in oxygen. A cytological analysis was then made of chromosomal rearrangements (interchanges and interstitial deletions) in microspores at the four- to five-day interval following irradiation. The frequency of aberrations was strikingly reduced when nitrogen replaced air and increased when oxygen replaced air. The magnitude of this effect is very great, the increase in aberration frequency in exposures made in oxygen as compared with nitrogen being as much as fivefold at higher doses. Aberration frequencies were also reduced in additional exposures made in other gases such as helium and argon, demonstrating that the availability of oxygen is a very important factor in the production of chromosomal rearrangements by x-rays. It is not possible on the present evidence to decide whether this effect of oxygen is exerted on the initial breakage mechanism or on the recovery process.

---

[*] This work was done under Contract No. W-7405-eng-26 for the Atomic Energy Commission, Oak Ridge, Tennessee.

[†] On leave from the University of Kentucky.

[1] Thoday, J. M., and Read, John, *Nature*, 160, 608 (1947); *Nature*, 163, 133 (1949).

[2] Hayden, B., and Smith, L., *Genetics*, 34, 26 (1949).

[3] Sax, Karl, *Ibid.*, 23, 494 (1938); 25, 41 (1940).

[4] Lea, D. E., *Actions of Radiations on Living Cells*, Cambridge Univ. Press, Cambridge, England, 1945.

[5] Catcheside, D. G., *Advances in Genetics*, 2, 271 (1948).

[6] Catcheside, D. G., Lea, D. E., and Thoday, J. M., *J. Genetics*, 47, 113 (1946).

[7] The authors would like to express their appreciation to Dr. W. A. Arnold of this laboratory for making these calculations.

# THE INFLUENCE OF THE Y-CHROMOSOME ON X-RAY INDUCED MUTATION RATE IN DROSOPHILA MELANOGASTER

BY JEAN KERSCHNER*

DEPARTMENT OF ZOOLOGY, UNIVERSITY OF PENNSYLVANIA, PHILADELPHIA,

Communicated by C. W. Metz, September 10, 1949

*Introduction.*—There is an increasing interest in the rôle played by heterochromatin in the synthesis of cellular proteins.[1] Heitz coined the term "heterochromatin" in 1928, and later[2] postulated from purely cytological evidence that heterochromatin would probably prove to be genetically inert. Subsequent work has borne out the essential correctness of this hypothesis. Although practically devoid of genes, heterochromatin, nevertheless, is not inactive.

There are several lines of evidence which indicate that heterochromatin plays a rôle in cellular metabolism. (1) Excessive amounts of heterochromatin may cause sterility, as in the case of the *XYYY* male of *Drosophila melanogaster*.[3] (2) The amount of heterochromatin in the nucleus influences the amount of cytoplasmic nucleic acid. Caspersson and Schultz[4] found that *XXY* ovaries of *D. melanogaster* are richer in cytoplasmic nucleic acid than *XX* ovaries. (3) Heterochromatin causes instability in certain genes upon their transposition from normal euchromatic sites to ones located in or near heterochromatin.[5] Paradoxically, some of these unstable genes are rendered more stable by the addition of extra heterochromatin in the form of supernumerary *Y*-chromosomes.[6] (4) The *Y*-chromosome of Drosophila is instrumental in increasing mutation rate. According to recent work of Mampell[7, 8] there are two actions of the *Y*: first, in *D. pseudoobscura* the *Y*-chromosome is necessary for the activity of the "mutator" gene, a recessive factor, located on chromosome 2, which causes increased mutability; second, the *Y*-chromosome of one species can apparently cause an increased mutability in flies of another species.

It is of interest, therefore, to determine the effect of heterochromatin on x-ray induced mutation frequency. The experiments described here were designed to test the effect of an extra *Y*-chromosome on x-ray induced visible and sex-linked recessive lethal mutation frequency in *D. melanogaster*.

*Materials and Methods.*—The stock used, No. 760 Stock from the Lankenau Hospital Research Institute,[9] is one in which the *X*-chromosome carries the white-mottled-4 ($w^{m4}$) inversion (breaks to the left of white, at band 3*C*2, and to the right of band 20, the bobbed locus), and in which the autosomes are from the Oregon-*R* wild stock.

The $w^{m4}$ mutant type, first described by Muller,[10] is considered one of the "eversporting" displacements. Presumably the normal allele of white, upon its transposition (by means of the $w^{m4}$ inversion) to heterochromatin, becomes unstable and mutates somatically in the developing eye disc to the recessive allele, thereby producing the characteristic mottled phenotype. An extra *Y*-chromosome increases the stability of $w^{m4}$ and the phenotype becomes more nearly normal.

In the experiments described here, use was made of this effect of an extra *Y*-chromosome. Phenotypically distinguishable *XY* and *XYY* males from the $w^{m4}$ stock were x-rayed simultaneously and their progeny examined for induced visible (attached-*X* method)[11] and sex-linked recessive lethal (*ClB* method)[12] mutations. Phenotypically distinguishable *XX* and *XXY* females from the same stock were also irradiated simultaneously and their progeny examined for visibles.[11]

In conducting the experiments on x-rayed males, the treated flies were bred singly (pair matings) and the males removed after three days to insure testing of only those sperm which were mature at the time of treatment. In the case of irradiated females, the treated flies were bred singly (pair matings) to males, from the attached-*X* stock, which carried the sex-linked recessive $fu^{o}$; eggs were collected over a ten-day period. For both the male and female series, parallel unirradiated cultures were run as controls.

In the experiments on irradiated males, only certain types of sperm were studied. The *XY* males from the $w^{m4}$ stock form *X*- and *Y*-bearing sperm in the expected ratio of 1:1. The *XYY* males form four kinds of sperm, *XY*-, *Y*-, *YY*- and *X*-bearing in the ratio of approximately 7:7:1:1 (this ratio indicates preferential synapsis of the two *Y*-chromosomes at meiosis: *YY* synapsis in approximately 76% of spermatocytes and *XY* synapsis in 24%). The *ClB* method, which was employed for detecting sex-linked recessive lethals, tests only *X*-bearing sperm from *XY* males and *X*- and *XY*-bearing sperm from *XYY* males. In these tests for induced lethals in sperm, it was not possible with the *ClB* stock employed to distinguish between *X*- and *XY*-bearing sperm of the *XYY* males. With the attached-*X* method for visibles, however, such a distinction could be made. In the data presented below, where the total visible rate was measured, both types of sperm from *XY* males and all four types from *XYY* males are included.

The *XXY* females from the $w^{m4}$ stock form four kinds of eggs: *XY*, *X*, *XX* and *Y*, in the approximate ratio of 9:9:1:1. In the irradiation studies on the *XXY* females, *XY*, *X* and *XX* eggs were scored.

*Results.*—On the basis of the experiments performed, three types of comparison can be made: first, a comparison of the frequency of visible mutations induced in the *X*- and *XY*-bearing sperm of *XYY* males and in the *X*-bearing sperm of both *XY* and *XYY* males; second, a comparison of the relative frequencies of visible mutations in the progeny of irradiated *XX* and *XXY* females; third, a comparison of the relative frequencies of total visible and sex-linked recessive lethal mutations in the progeny of irradiated $XY$ and $XYY$ males.

In terms of the actual data, these three comparisons are as follows: (1) The $XY$ sperm of treated $XYY$ males show a lower visible mutation rate than $X$-bearing sperm from the same males (22 mutants[18] in 2974 flies, or 0.74% as opposed to 9 mutants in 521 flies, or 1.73%; the exact $P$ value of this difference, as determined by chi square, is 0.062—see table 1). The $X$-bearing sperm from the two kinds of males show mutation rates which are not significantly different (36 mutants in 1452 flies, or 2.47%, and 9 mutants in 521 flies, or 1.73%; the $P$ value of this difference is 0.434—see table 1). (2) There is no significant difference in the visible mutation rates in x-rayed $XXY$ and $XX$ females (23 mutants in 4679 flies, or 0.49%, as opposed to 12 mutants in 2615 flies, or 0.46%; the $P$ value of this difference is 0.996—see table 2). (3) The total visible mutation rate for irradiated $XYY$ males is significantly lower than that for

TABLE 1

VISIBLE MUTATIONS SCORED ACCORDING TO TYPE OF SPERM

| | TOTAL FLIES | NO. OF MUTANTS | MUTATION RATE, % | $P$ VALUE OF DIFFERENCE |
|---|---|---|---|---|
| | *Irradiated*[21] *(4000 r)* | | | |
| $X$ sperm from $XY$ males | 1452 | 36 | 2.47 | 0.434 |
| $X$ sperm from $XYY$ males | 521 | 9 | 1.73 | 0.062 |
| $XY$ sperm from $XYY$ males | 2974 | 22 | 0.74 | |
| | *Control (Untreated)* | | | |
| $X$ sperm from $XY$ males | 1176 | 4 | 0.34 | 0.962 |
| $X$ sperm from $XYY$ males | 243 | 0 | 0.00 | 0.77 |
| $XY$ sperm from $XYY$ males | 1758 | 2 | 0.11 | |

TABLE 2

VISIBLE MUTATION FREQUENCY IN FEMALES

| | NO. OF CULTURES | TOTAL FLIES | FLIES PER CULTURE | NO. OF MUTANTS | MUTATION RATE, % | $P$ VALUE DIFFERENCE |
|---|---|---|---|---|---|---|
| | | *Irradiated (2000 r)* | | | | |
| $XX$ females | 123 | 2615 | 21.2 | 12 | 0.46 | 0.996 |
| $XXY$ females | 270 | 4679 | 17.3 | 23 | 0.49 | |
| | | *Control (Untreated)* | | | | |
| $XX$ females | 41 | 1660 | 40.5 | 0 | 0.00 | |
| $XXY$ females | 42 | 1325 | 31.5 | 0 | 0.00 | |

*XY* males (32 mutants in 6498 flies, or 0.49%, as opposed to 37 mutants in 3151 flies, or 1.17%, the difference having a *P* value of 0.0005—see table 3). And, in the case of the sex-linked recessive lethals,[14] where the observations are completely objective, the mutation rate obtained for the *XYY* males is much lower than the rate obtained for *XY* males (22 cultures out of a total of 344 showed lethals, giving a rate of 6.40%, as opposed to 35 cultures out of a total of 286, giving a rate of 12.24%; the *P* value of the difference is 0.014—see table 4).

*Discussion.*—In connection with the experiments presented here, there are three points which bear discussion.

TABLE 3

VISIBLE MUTATION FREQUENCY IN MALES

| | NO. OF CULTURES | TOTAL FLIES | FLIES PER CULTURE | NO. OF MUTANTS | MUTATION RATE, % | *P* VALUE DIFFERENCE |
|---|---|---|---|---|---|---|
| | | | *Irradiated (4000 r)* | | | |
| *XY* males | 421 | 3151 | 7.48 | 37 | 1.17 | 0.0005 |
| *XYY* males | 655 | 6498 | 9.77 | 32 | 0.49 | |
| | | | *Control (Untreated)* | | | |
| *XY* males | 48 | 2180 | 45.4 | 3 | 0.14 | 0.065 |
| *XYY* males | 61 | 3145 | 51.6 | 2 | 0.064 | |

TABLE 4

RECESSIVE SEX-LINKED LETHAL MUTATION FREQUENCY IN MALES

| | TOTAL $F_2$ CULTURES | NO. OF CULTURES WITH LETHAL | MUTATION RATE, % | *P* VALUE OF DIFFERENCE |
|---|---|---|---|---|
| | | *Irradiated (4000 r)* | | |
| *XY* males | 286 | 35 | 12.24 | 0.014 |
| *XYY* males | 344 | 22 | 6.40 | |
| | | *Control (Untreated)* | | |
| *XY* males | 347 | 2 | 0.58 | 0.085 |
| *XYY* males | 189 | 2 | 1.06 | |

In the first place, it is demonstrated that an extra *Y*-chromosome, when present in the male, significantly decreases both the visible and sex-linked lethal mutation rates. Moreover, there are two lines of evidence which indicate that the influence of the extra *Y* is not exerted until after meiosis: (1) the visible mutation rate induced in *X*-bearing sperm of *XY* males is not significantly different from the rate induced in *X*-bearing sperm of *XYY* males; (2) *XY*-bearing sperm from *XYY* males show a lower frequency of induced visibles than *X*-bearing sperm from the same males. It is not possible in the sex-linked recessive lethal series to determine whether the action of the extra $Y$ precedes or follows meiosis, since the $X$- and $XY$-bearing sperm of the $XYY$ males are indistinguishable.

Second, the exact mode of action of the extra $Y$ in the sperm is not understood. The extra heterochromatin may shield the chromosomes from the ionizing effects of the x-rays, or it may provide a reservoir of material for chromosome restitution, thereby decreasing the frequency of chromosomal rearrangements and consequently decreasing the frequency of position effects. If the latter were true, the lack of effect of the $Y$ in treated females might be understandable, since interchromosomal rearrangements are known to occur at a negligible rate in irradiated eggs.[15–17] Recent preliminary studies of Dr. H. B. Glass seem to indicate that the frequency of reciprocal translocations induced by x-rays in Drosophila eggs can be increased if the females are exposed to infra-red following x-ray treatment (personal communication). This work suggests a type of experiment which may throw light on the mode of action of the $Y$-chromosome in the present experiments. If an extra $Y$ decreases the frequency of interchromosomal exchange, infra-red post-treatment of irradiated females should provide experimental conditions favorable for the detection of such action. Additional experiments which may help to throw light on the failure of the $Y$ action in females are planned. They will involve a comparative study on sex-linked recessive lethal mutations induced in $XX$ and $XXY$ females.

Finally, if the extra $Y$ chromosome retards maturation so that there are more immature germ cells in $XYY$ testes than in $XY$ testes, this might conceivably account for the lowered mutation rate.[18–20] Experiments to test this possibility are already under way.

*Summary.*—The results of experiments described here show that an extra $Y$-chromosome, when present in the male of *Drosophila melanogaster*, decreases x-ray induced visible and sex-linked lethal mutation frequency. In the female, an extra $Y$ appears to be without effect. The possible mode of action of the $Y$ has been discussed and future lines of investigation proposed.

---

[*] This work was done during the tenure of Predoctorate Fellowships from the Donner Foundation and from the U. S. Public Health Service (National Cancer Institute).

[1] Wright, S., *Am. Nat.*, **79**, 289–303 (1945).

[2] Heitz, E., *Planta. Archiv für wissenschaftliche Botanik. Abteilung E der Zeitschrift für Biologie*, **18**, 571–636 (1932).

[3] Schultz, J., *Proc. VIIth Int. Genet. Cong., Edinburgh*, pp. 257–262 (1941).

[4] Caspersson, T., and Schultz, J., *Nature*, **142**, 294–295 (1938).

[5] Demerec, M., *Genetics*, **25**, 618–627 (1940).

[6] Gowen, J. W., and Gay, E. H., these PROCEEDINGS, **19**, 122–126 (1933).

[7] Mampell, K., *Genetics*, **30**, 498–505 (1945).

[8] Mampell, K., *Ibid.*, **31**, 589–597 (1946).

[9] The author is indebted to Dr. Jack Schultz for this stock.

[10] Muller, H. J., *J. Genet.*, **22**, 299–334 (1930).

[11] No attempt was made to detect autosomal recessives.

[12] The $F_1$ progenies derived from $XYY$ males and $C1B$ females were examined microscopically in order to eliminate the possibility that patroclinous sons resulting from secondary non-disjunction might mask a lethal.

[13] Breeding tests were conducted on all flies showing conspicuous, symmetrical alterations in phenotype, and such changes were scored as mutants even in those cases where inviability or sterility prevented actual progeny counts.

[14] It should be noted that it was not possible to compare the recessive lethal rates in the two kinds of sperm ($X$ and $XY$) from $XYY$ males, as was possible in the case of visibles.

[15] Glass, H. B., *Genetics*, **25**, 117 (1940).

[16] Metz, C. W., and Bozeman, M. L., *Carnegie Inst. Wash. Year Book*, **41**, 237–242 (1942).

[17] Reynolds, J. P., these PROCEEDINGS, **27**, 204–208 (1941).

[18] Harris, B. B., *J. Hered.*, **20**, 299–302 (1929).

[19] Kossikov, K. V., *Genetics*, **22**, 213–224 (1937).

[20] Cooper, K. W., *J. Morph.*, **84**, 81–121 (1949).

[21] X-raying was done at the University of Pennsylvania Hospital, through the courtesy of the Department of Radiology. A medium-voltage therapy machine was used with no filter; target distance, 17 cm.; average output, 450 r per minute; current maintained at 8 milliamperes and 135 kilovolts.

---

# ON A THEOREM OF WEYL CONCERNING EIGENVALUES OF LINEAR TRANSFORMATIONS. I*

BY KY FAN

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF NOTRE DAME

Communicated by Hermann Weyl, September 22, 1949

H. Weyl[1] recently proved the following theorem:

THEOREM. *Let $A$ be a linear transformation in the n-dimensional unitary space $C_n$. Let the eigenvalues of $A$ and $A^*A$ be denoted by $\lambda_i$ and $\kappa_i (1 \leq i \leq n)$, respectively, which are so arranged that*

$$|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_n|, \qquad \kappa_1 \geq \kappa_2 \geq \ldots \geq \kappa_n. \qquad (1)$$

*For any non-decreasing function $\omega(t)$ on $t > 0$ such that $\omega(e^t)$ is a convex function of t and $\omega(0) = \lim_{t \to 0} \omega(t) = 0$, $\lambda_i$ and $\kappa_i$ satisfy the inequalities:*

$$\sum_{i=1}^{q} \omega(|\lambda_i|^2) \leq \sum_{i=1}^{q} \omega(\kappa_i) \qquad (1 \leq q \leq n). \qquad (2)$$

In the present note, we prove three related theorems. Theorem 1 gives an extremum property of the sum of the first $q$ eigenvalues for Hermitian transformations. This property furnishes a recurrent characterization of successive eigenvalues without referring to any eigenvector. Theorem 2 gives a similar but stronger property for all normal transformations. For an arbitrary linear transformation $A$ and for a positive integer $s$, we have in Theorem 3 inequalities comparing the eigenvalues of $(A^s)^*A^s$ with those of $A^*A$. Finally we shall see that Weyl's theorem in the most important case $\omega(t) = t^s$ $(s = 1, 2, 3, \ldots)$ can be derived from Theorems 1 and 3. The general case of Weyl's theorem will be discussed in a forthcoming note.[2] All linear transformations considered here are assumed to be in the $n$-dimensional unitary space $C_n$, but the results can be easily carried over to completely continuous linear operators in Hilbert space, especially to continuous kernels of linear integral equations.

THEOREM 1. *Let the eigenvalues* $\lambda_i$ *of a Hermitian transformation* $H$ *be so arranged that* $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. *For any positive integer* $q \leq n$, *the sums* $\sum_{i=1}^{q} \lambda_i$ *and* $\sum_{i=1}^{q} \lambda_{n+1-i}$ *are, respectively, the maximum and minimum of* $\sum_{j=1}^{q} (Hx_j, x_j)$, *when q orthonormal vectors* $x_j$ $(1 \leq j \leq q)$ *vary in the space.*[3]

*Proof:* Let $\varphi_i (1 \leq i \leq n)$ be an orthonormal set of eigenvectors of $H$: $H\varphi_i = \lambda_i\varphi_i$. For each $j$, we write

$$(Hx_j, x_j) = \lambda_q \sum_{i=1}^{n} |(x_j, \varphi_i)|^2 + \sum_{i=1}^{q} (\lambda_i - \lambda_q) |(x_j, \varphi_i)|^2 + \sum_{i=q+1}^{n} (\lambda_i - \lambda_q) |(x_j, \varphi_i)|^2. \quad (3)$$

If $\|x_j\| = 1$, then

$$(Hx_j, x_j) \leq \lambda_q + \sum_{i=1}^{q} (\lambda_i - \lambda_q) |(x_j, \varphi_i)|^2$$

and therefore

$$\sum_{i=1}^{q} \lambda_i - \sum_{j=1}^{q} (Hx_j, x_j) \geq \sum_{i=1}^{q} (\lambda_i - \lambda_q) \{1 - \sum_{j=1}^{q} |(x_j, \varphi_i)|^2\}. \quad (4)$$

If $x_j$ $(1 \leq j \leq q)$ are orthonormal, then $\sum_{j=1}^{q} |(x_j, \varphi_i)|^2 \leq \|\varphi_i\|^2 = 1$, so that the right-hand side of (4) is $\geq 0$. But the left-hand side vanishes for $x_j = \varphi_j$ $(1 \leq j \leq q)$. This proves the maximum property.

THEOREM 2. *Let* $\lambda_i$ *be the eigenvalues of a normal transformation* $N$ *so arranged that* $|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_n|$. *Let s, q be two positive integers* $(q \leq n)$. *Then* $\sum_{i=1}^{q} |\lambda_i|^{2s}$ *is the maximum of* $\sum_{j=1}^{q} \|(UN)^s x_j\|^2$, *when U runs over all unitary transformations and* $x_j$ $(1 \leq j \leq q)$ *runs over all sets of q orthonormal vectors in* $C_n$.

*Proof:* We need only prove the inequality

$$\sum_{j=1}^{q} \|(UN)^s x_j\|^2 \leq \sum_{i=1}^{q} |\lambda_i|^{2s}. \tag{5}$$

As $\|UNx_j\|^2 = (N^*Nx_j, x_j)$, the case $s = 1$ of (5) follows from Theorem 1. We proceed by induction, assuming that (5) is true for $s$. Let $\varphi_i (1 \leq i \leq n)$ be an orthonormal set of eigenvectors of $N$: $N\varphi_i = \lambda_i \varphi_i$. Consider a unitary transformation $U$ and $q$ orthonormal vectors $x_j$ $(1 \leq j \leq q)$. We have $\|(UN)^{s+1}x_j\|^2 = \sum_{i=1}^{n} |\lambda_i|^2 \cdot |((UN)^s x_j, \varphi_i)|^2$. If we split this sum into three parts in a way similar to (3), we see that for each $j$:

$$\|(UN)^{s+1}x_j\|^2 \leq |\lambda_q|^2 \cdot \|(UN)^s x_j\|^2 + \sum_{i=1}^{q} (|\lambda_i|^2 - |\lambda_q|^2) \cdot |((UN)^s x_j, \varphi_i)|^2. \tag{6}$$

As $x_j (1 \leq j \leq q)$ are orthonormal, we have for each $i$:

$$\sum_{j=1}^{q} |((UN)^s x_j, \varphi_i)|^2 \leq \|(N^*U^*)^s \varphi_i\|^2 = \|(U^*N^*)^s U^* \varphi_i\|^2. \tag{7}$$

Using first (6), then (7) and our assumption of induction (i.e. (5) is true for $s$), we get

$$\sum_{i=1}^{q} |\lambda_i|^{2s+2} - \sum_{j=1}^{q} \|(UN)^{s+1}x_j\|^2 \geq \sum_{i=1}^{q} (|\lambda_i|^2 - |\lambda_q|^2)[|\lambda_i|^{2s} - \|(U^*N^*)^s U^* \varphi_i\|^2]. \tag{8}$$

Denote by $d_q$ the expression on the right-hand side of (8), we have

$$d_{q+1} - d_q = (|\lambda_q|^2 - |\lambda_{q+1}|^2)[\sum_{i=1}^{q} |\lambda_i|^{2s} - \sum_{i=1}^{q} \|(U^*N^*)^s U^* \varphi_i\|^2]. \tag{9}$$

As $U^* \varphi_i (1 \leq i \leq q)$ are orthonormal, our assumption of induction shows that the right-hand side of (9) is $\geq 0$, and $d_{q+1} \geq d_q$. But $d_1 = 0$, hence $d_q \geq 0$. This proves that (5) is also true for $s + 1$.

THEOREM 3. *Let $A$ be an arbitrary linear transformation and $s$ be any positive integer. Let the eigenvalues of $(A^s)^*A^s$ be denoted by $\kappa_i^{(s)}$ $= \kappa_i)$ and so arranged that $\kappa_1^{(s)} \geq \kappa_2^{(s)} \geq \ldots \geq \kappa_n^{(s)}$. Then for any positive integer $q \leq n$, we have*

$$\sum_{i=1}^{q} \kappa_i^{(s)} \leq \sum_{i=1}^{q} \kappa_i^s. \tag{10}$$

*Proof:* Let $A = UH$ be the polar decomposition of $A$, where $U$ is unitary and $H$ is the non-negative square root of $A^*A$. The eigenvalues of

$H$ are $\kappa_i^{1/s}$. By Theorem 2, any $q$ orthonormal vectors $x_j (1 \leq j \leq q)$ satisfy

$$\sum_{j=1}^{q} ((A^s)^* A^s x_j, x_j) = \sum_{j=1}^{q} \|(UH)^s x_j\|^2 \leq \sum_{i=1}^{q} \kappa_i^s. \qquad (11)$$

But by Theorem 1, $\sum_{i=1}^{q} \kappa_i^{(s)}$ is the maximum of the first $\sum$ sum in (11), when the $q$ orthonormal vectors $x_j$ vary. Thus (10) is proved.

We now prove the case $\omega(t) = t^s (s = 1, 2, 3, \ldots)$ of Weyl's theorem. Here we use the same notation as in the theorem stated at the beginning of this note. Using Schur's superdiagonal form of matrices, it is clear that there exist $n$ orthonormal vectors $y_i (1 \leq i \leq n)$ such that $|\lambda_i|^2 \leq \|Ay_i\|^2 (1 \leq i \leq n)$ and therefore $\sum_{i=1}^{q} |\lambda_i|^2 \leq \sum_{i=1}^{q} \|Ay_i\|^2$. But applying Theorem 1 to $A^*A$, we find $\sum_{i=1}^{q} \|Ay_i\|^2 \leq \sum_{i=1}^{q} \kappa_i$. Hence

$$\sum_{i=1}^{q} |\lambda_i|^2 \leq \sum_{i=1}^{q} \kappa_i. \qquad (12)$$

As in Theorem 3, we denote by $\kappa_i^{(s)}$ the eigenvalues of $(A^s)^*A^s$ arranged in descending order (in particular, $\kappa_i^{(1)} = \kappa_i$). Applying (12) to the transformation $A^s$, we get

$$\sum_{i=1}^{q} |\lambda_i|^{2s} \leq \sum_{i=1}^{q} \kappa_i^{(s)},$$

which together with (10) gives the case $\omega(t) = t^s$ of (2).

---

[*] This work was supported in part by the Office of Naval Research.

[1] Weyl, H., "Inequalities between the Two Kinds of Eigenvalues of a Linear Transformation," these PROCEEDINGS, **35**, 408–411 (1949).

[2] Fan, K., "On a Theorem of Weyl concerning Eigenvalues of Linear Transformations. II," to be published in these PROCEEDINGS.

[3] An alternative form of Theorem 1: $\sum_{i=1}^{q}\lambda_i$ and $\sum_{i=1}^{q}\lambda_{n+1-i}$ are, respectively, the maximum and minimum of the trace of $HP$, when $P$ runs over all projections on $q$-dimensional linear subspaces. There is also a similar alternative form of Theorem 2.

# ON SELF-ADJOINT DIFFERENTIAL EQUATIONS OF SECOND ORDER

By Walter Leighton

Department of Mathematics, Washington University

Communicated by Marston Morse, September 21, 1949

This note is an abstract of a longer paper which will be published elsewhere.

The paper is concerned with the behavior near $x = +\infty$ of solutions of the self-adjoint differential equation

$$(r(x)y')' + p(x)y = 0, \tag{1}$$

where $r(x) > 0$ and $r(x)$, $r'(x)$ and $p(x)$ are continuous functions of the real variable $x$ for $x > 0$. The theorem which follows enables one to compare the rapidity of oscillation of a solution of (1) with that of a solution of

$$(r(x)z')' + p_1(x)z = 0. \tag{2}$$

THEOREM 1. *If for all solutions $y(x)$ and $z(x)$ of (1) and (2), respectively, the function*

$$W(x) = r(x)[y(x)z'(x) - y'(x)z(x)]$$

*either is eventually of fixed sign or has the limit zero as $x$ becomes infinite, the zeros of some solution $y(x)$ eventually separate those of each non null solution $z(x)$, and vice versa.*

COROLLARY. *If all solutions $y(x)$ and $z(x)$ are bounded and if the limit*

$$\lim_{x \to \infty} \int_a^x |p - p_1| dx$$

*exists (in the strict sense), the conclusion of Theorem 1 is valid.*

The results stated are valid whether or not the solutions of (1) and of (2) are oscillatory near $x = \infty$.

In studying (1) it is frequently useful to write linearly independent solutions of this equation in the *canonical* form

$$u(x) \sin v(x), \qquad u(x) \cos v(x) \tag{3}$$

used by Milne, Fatou and others. Here $u(x)$ and $v(x)$ may be assumed to be positive for $x > 0$. A companion theorem of an earlier result[1] is this.

THEOREM 2. *If $p(x) > 0$ and $p(x)r(x)$ is monotone decreasing as $x$ increases, there exists a positive constant $c^2$ such that $u(x) \geq c^2$ for $x \geq a > 0$.*

If $r = 1$, and $p(x) > 0$ and of class $C^2$, write $p(x) = [q(x)]^{-4}$. The following theorem may then be stated.

THEOREM 3. *Suppose $Q = (1 + q^3q'')$ is positive for $x$ large. If $Q$ is monotone increasing as $x$ increases, each solution $y(x)$ of (1) has the property* *that* $|y(x)| \leq Mq(x)$, *where* $M$ *is a positive constant. If* $Q$ *decreases monotonely as* $x$ *increases, corresponding to each solution* $y(x)$ *there exists a positive constant* $m$ *such that*

$$u(x) \geq mq(x),$$

*where* $u(x)$ *is any coefficient in* (*3*).

A number of necessary conditions that solutions of (1) be oscillatory (i.e., sufficient conditions for non-oscillation) near $x = \infty$ are given. A principal theorem is the following.

THEOREM 4. *If* $r = 1$ *and* $x^2p(x) - {}^1/_4$ *is a positive monotone function of* $x$, *a necessary condition that a solution* $y(x)$ *of* (*1*) *be oscillatory near* $x = \infty$ *is that*

$$\lim_{x \to \infty} \int_1^x [(x^2p(x) - {}^1/_4)^{1/2}/x]dx = \infty.$$

A number of corollaries may be stated. The following is of interest as a companion result to an earlier one given by the present author.[2]

COROLLARY. *If* $p(x)r(x)$ *is a positive monotone function of* $x$ *for* $x$ *large, a necessary condition that solutions* $y(x)$ *of* (*1*) *be oscillatory neighboring* $x = \infty$ *is that not both limits*

$$\lim_{x \to \infty} \int_1^x \frac{dx}{r(x)}, \qquad \lim_{x \to \infty} \int_1^x p(x)dx$$

*exist.*

Proofs of the theorems stated and of other related results, as well as a more complete bibliography, will be given in the longer paper of which this is an abstract.

---

[1] These PROCEEDINGS, **35**, 190 (1949).

[2] *Ibid.*, **35**, 193 (1949).

---

# THE COEFFICIENT PROBLEM FOR BOUNDED SCHLICHT FUNCTIONS*

BY H. L. ROYDEN†

STANFORD UNIVERSITY

Communicated by S. Lefschetz, September 9, 1949

The theory of the class of bounded schlicht functions is in some respects more symmetric than that of the general class of schlicht functions, and it is the purpose of the present paper to give some of the preliminary theory and to indicate some of the problems attached to this class of functions.

1. Consider the two classes of functions: the class $\mathcal{S}$, consisting of all functions of the form

$$w = f(z) = z + a_2z^2 + a_3z^3 + \ldots$$

which are regular and schlicht in $|z| < 1$; the class $\mathcal{B}\mathcal{S}$, consisting of all functions of the form

$$w = f(z) = b_1z + b_2z^2 + b_3z^3 + \ldots$$

with $b_1 > 0$, which are regular, schlicht and bounded by unity in $|z| < 1$. The set of all the points $(a_2, \ldots, a_n)$ which belong to functions of class $\mathcal{S}$ form, in $(2n - 2)$-dimensional space, the closed domain $V_n$, which has been studied by Schaeffer and Spencer.[6] Similarly, the set of all points $(b_1, b_2, \ldots, b_n)$ belonging to functions of class $\mathcal{B}\mathcal{S}$ form a region $\mathcal{V}_n$ in $(2n - 1)$-dimensional space. Each bounded function of class $\mathcal{S}$ is transformed into a function of $\mathcal{B}\mathcal{S}$ on division by its maximum modulus, and each function of $\mathcal{B}\mathcal{S}$ is transformed into a bounded function of $\mathcal{S}$ on division by $b_1$.

Since the class $\mathcal{B}\mathcal{S}$ becomes compact with the addition of the function $w = 0$, the region $\mathcal{V}_n + 0$ (the origin) is closed. It can be shown that a necessary and sufficient condition that the point $(b_1, b_2, \ldots, b_n)$ be an interior point of $\mathcal{V}_n$ is the existence of a function $f(z) \in \mathcal{B}\mathcal{S}$ belonging to the point and bounded by $r < 1$ in $|z| < 1$. Hence $\mathcal{V}_n + 0$ is the closure of its interior.

Each interior point $(a_2, \ldots, a_n)$ of $V_n$ determines a segment $(b_1, b_1a_2, \ldots, b_1a_n)$ of $\mathcal{V}_n$ as $b_1$ varies from 0 to $b_1^*$, the maximum value $b_1$ can have for the point $(b_1, b_1a_2, \ldots, b_1a_n)$ to belong to $\mathcal{V}_n$. The end points can be the only points of intersection with the boundary of $\mathcal{V}_n + 0$, since all the other points are interior points. Now $b_1^*$ will be the reciprocal of the minimum of the maximum moduli of the functions of class $\mathcal{S}$ belonging to the point $(a_2, \ldots, a_n)$. We have $b_1^* > 0$, since the interior points of $V_n$ possess bounded functions.[6]

To construct $\mathcal{V}_n$ starting from $V_n$, one first constructs a cone in $(2n - 1)$-dimensional space with vertex at the origin and base $V_n$ in the hyperplane $b_1 = 1$. Cut each ray from the origin to an interior point of $V_n$ off at a vertical distance $b_1^*$ from the origin. The segments from the origin to $(b_1^*, b_1^*a_2, \ldots, b_1^*a_n)$ form $\mathcal{V}_n$.

It now follows that $b_1^*$ is a continuous function of the point $(a_2, \ldots, a_n)$, for if $b_1^*$ had more than one limit point as we approach the point $(a_2, \ldots, a_n)$ we would have more than one boundary point of $\mathcal{V}_n$ on the segment $(b_1, b_1a_2, \ldots, b_1a_n)$, $0 < b_1 \leqslant b_1^*$, which is impossible. From this follows the continuity of the minimum of the maximum moduli of the functions of $\mathcal{S}$ belonging to interior points of $V_n$.

Since $V_n$ is topologically equivalent to the closed $(2n - 2)$-dimensional sphere, $\mathcal{V}_n + 0$ is equivalent to the closed $(2n - 1)$-dimensional sphere.

2. Let the continuous, real valued function $F$ have a continuous and non-vanishing gradient in an open set containing $\mathcal{V}_n + 0$. Then on the basis of the formula (2.5.2) of reference 6, one has the following lemma, which has also been obtained by Charzyński.[2]

LEMMA: *Every function $w = f(z)$ of class $\mathcal{BS}$ belonging to a point $(b_1, b_2, \ldots, b_n)$ where $F$ attains its maximum in $\mathcal{V}_n$ must satisfy the differential equation*

$$\left(\frac{z}{w}\frac{dw}{dz}\right)^2 R(w) = Q(z) \qquad (1)$$

*where*

$$R(w) = \sum_{\nu = -n+1}^{n-1} \frac{\beta_\nu}{w^\nu},\ Q(z) = \sum_{\nu = -n+1}^{n-1} \frac{B_\nu}{z^\nu}. \qquad (2)$$

*Moreover, the functions $R(w)$ and $Q(z)$ are real and non-negative on the unit circumference, and each has at least one zero there. If we set $F_k = \frac{\partial F}{\partial b_k}$, then*

$$\begin{aligned} \beta_\nu &= \sum_{k=\nu+1}^{n} F_k b_k^{(\nu+1)}, & \beta_{-\nu} &= \bar{\beta}_\nu \\ & & \beta_0 &= {}^1/_2 F_1 b_1 + Re \sum_{k=2}^{n} b_k F_k \qquad (3) \\ B_\nu &= \sum_{k=1}^{n-\nu} k b_k F_{k+\nu}, & B_{-\nu} &= \bar{B}_\nu \\ & & B_0 &= {}^1/_2 F_1 b_1 + Re \sum_{k=2}^{n} k b_k F_k. \end{aligned}$$

Thus it is clear that an equation of the form (1) is satisfied by the functions belonging to a dense set of boundary points, and by a limiting process such an equation is seen to be satisfied at every boundary point (except of course at $(0, \ldots, 0)$), where now the $F_k$ are merely certain constants. Such equations will be called $\mathfrak{D}_n$-equations, provided that if $k$ is the largest integer such that $B_k \neq 0$, then $b_1^k \beta_k = B_k$ and all $\beta\nu = 0$ for $\nu > k$. This last condition is obviously necessary for the existence of solutions analytic at $z = 0$. A function $w = f(z)$ which is regular in $|z| < 1$ and which has a positive derivative at the origin will be called a $\mathfrak{D}_n$-function if it satisfies a $\mathfrak{D}_n$-equation. We must have $b_1 < 1$ for $f(z)$ to be bounded by unity, but this is the only restriction to be placed on $b_1$.

If $\zeta = \int \sqrt{Q(z)}\, dz/z$, we let $\Gamma_z$ be the set of arcs $Re(\zeta) =$ constant which have one or both end-points at a zero of $Q(z)$. This $\Gamma_z$-structure has been investigated in [6]. The $\Gamma_w$-structure is similarly defined in connection with $R(w)$, and in our case it has exactly the same form as the $\Gamma_z$-structure. Using methods similar to those in reference 6, the following theorem can be established.

THEOREM I: *If $w = f(z)$ is a $\mathfrak{D}_n$-function, then it is schlicht in $|z| < 1$ and maps $|z| < 1$ onto the interior of the unit circle minus a subcontinuum of $\Gamma_w$ containing the circle $|w| = 1$. By continuation from within $|z| < 1$ the function $f(z)$ is regular on $|z| = 1$ with the exception of finitely many points where it has the local development*

$$f(z) = \sum_{\nu = 0}^{\infty} \alpha_\nu (z - e^{i\varphi})^{\nu/m},$$

*$m$ a positive integer.*

Let $C_w$ be a subcontinuum of $\Gamma_w$ interior to and containing the circle $|w| = 1$. Let

$$\psi(w) = w + \alpha w^n + \beta w^{n+1} + \ldots$$

be regular, schlicht and bounded by unity in the complement of $C_w$ in $|w| < 1$. Then, using the method of Teichmüller,[7] we can show that $Re(\alpha\beta_{n-1}) \leqslant 0$, with equality if and only if $\psi(w) = w$. From this follows Theorem II.

THEOREM II: *Every $\mathfrak{D}_n$-function belongs to some boundary point of $\mathfrak{V}_n$, and to any given boundary point of $\mathfrak{V}_n$ (except $(0, 0, \ldots, 0)$) there belongs one and only one $\mathfrak{D}_n$-function.*

THEOREM III: *Of all the functions of class $\mathfrak{S}$ which belong to the interior point $(a_2, \ldots, a_n)$ of $V_n$, the one which has the smallest maximum modulus is unique and is $1/b_1^*$ times the $\mathfrak{D}_n$-function belonging to the point $(b_1^*, b_1^*a_2, \ldots, b_1^*a_n)$ of the boundary of $\mathfrak{V}_n$.*

This is a modification of a problem considered by Carathéodory and Fejer,[1] who did not require their functions to be schlicht.

3. If a part of the boundary of $\mathfrak{V}_n$ is smooth, then we can choose the function $F$ of our lemma in such a manner that it vanishes on any given open subset $S$ of this part of the boundary and is non-positive on the remainder of the boundary. Then by the lemma equation (1) is satisfied at each point of $S$ by the function belonging to this point. Hence at each point of $S$ there is a complex number $\bar{\kappa}$ of unit modulus such that

$$Q(\bar{\kappa}) = 0, \qquad Q'(\bar{\kappa}) = 0. \tag{4}$$

If we assume $B_{n-1} \neq 0$, a necessary and sufficient condition for the existence of a $\bar{\kappa}$ satisfying equations (4) is the vanishing of their resultant. The $B_\nu$ are given by (3), where the $F_\nu$ now determine the normal direction to the boundary of $\mathfrak{V}_n$. Thus the resultant becomes a non-linear first order partial differential equation for $S$.

If we consider the second of equations (4) as determining $\bar{\kappa}(B_\nu, \bar{B}_\nu)$, then the resultant is equivalent to

$$2Q(\bar{\kappa}) = 0. \tag{5}$$

The characteristics of (5) are given by[8]

$$\frac{db_k}{dt} = 2\frac{\partial Q(\kappa)}{\partial F_k} = 2\sum_{\nu=-n+1}^{n-1} \kappa^\nu \frac{\partial B}{\partial F_k} + 2Q'(\kappa)\frac{\partial \kappa}{\partial F_k}.$$

Since $Q'(\kappa) = 0$, this results in

$$\frac{db_k}{dt} = kb_k + 2\sum_{\nu=1}^{k-1}(k-v)b_{k-\nu}\kappa^\nu. \qquad (6)$$

These are the same as the equations obtained by equating coefficients in the differential equation

$$\frac{\partial g(z, t)}{\partial t} = z\frac{\partial g(z, t)}{\partial z}\frac{1+\kappa z}{1-\kappa z}, \qquad (7)$$

which was obtained by Löwner[3] and shown to apply to arbitrary slit mappings. In our case the function $g(z, t)$ maps onto the complement of a subcontinuum of a fixed $\Gamma_w$-structure. With increasing $t$ the subcontinuum is cut away until a time $t_0$ when $g(z, t_0) = z$. With decreasing $t$ the subcontinuum grows and $g(z, t) \to 0$ as $t \to -\infty$. Thus the points $(1, 0, \ldots, 0)$ and $(0, 0, \ldots, 0)$ are critical points of the differential equation (5).

Had we started with the function $R(w)$ instead of $Q(z)$ we should have obtained a different equation whose characteristics are

$$\frac{db_k}{d_t} = -(b_k + 2\sum_{v=1}^{k-1}\kappa^v b_k^{(v+1)}).$$

These correspond to the dual Löwner equation

$$\frac{\partial f(z, t)}{\partial t} = -f\frac{1+\kappa f}{1+\kappa f}.$$

4. We conclude by giving the form of the body $\mathfrak{V}_2$. Here $Q(z)$ must be of the form $c(e^{i\theta}z^{-1} + 2 + e^{-i\theta}z)$ and $R(w)$ must be $b_1c(e^{i\theta}w^{-1} + 2 + e^{-i\theta}w)$. Equation (1) becomes

$$b_1^{1/2}(w + e^{i\theta})\frac{dw}{\sqrt{w^3}} = (z + e^{i\theta})\frac{dz}{\sqrt{z^3}}.$$

Hence by integration

$$\frac{b_1 z}{(1 - ze^{i\theta})^2} = \frac{w}{(1 - we^{i\theta})^2}. \qquad (8)$$

The functions $w = f(z)$ defined by (8) map the unit circle onto the unit circle minus a straight slit orthogonal to the unit circle at $w = -e^{-i\theta}$. Calculating the coefficient of $b_2$ we get $b_2 = -2(b_1 - b_1^2)\,e^{-i\theta}$, or $b_2 =$

$2b_1(1 - b_1)$. This is an equation for the boundary of $\mathcal{V}_2$ and is in accord with the result of Pick.[4]

It is interesting to note that for $n = 2$ the resultant of equations (4) becomes $B_0^2 = 4|B_1|^2$. Writing $b_2 = re^{i\varphi}$, we have

$$b_1^2(F_r^2 + F_\varphi^2) = ({}^1/_2 b_1 F_{b_1} + rF_r)^2. \tag{9}$$

Since we can change the argument of $b_2$ without changing its modulus merely by rotating both the $z$ and $w$ planes, $F$ must be independent of $\varphi$. Thus (9) becomes an ordinary differential equation which separates and integrates into

$$-\frac{1}{b_1} + c = -\frac{1}{2}\left(\frac{r}{b_1^2}\right).$$

Since $r = 0$ at $b_1 = 1$, $c$ must be zero giving $r = 2(b_1 - b_1^2)$ as before.

[*] This paper was written while the author was engaged on a project sponsored by the Office of Naval Research (Contract N6-ori, 154, Task III).

[†] Author's present address, Department of Mathematics, Harvard University.

[1] Carathéodory, C., and Fejer, L., *Rend. Circ. Mat. Palermo*, **12,** 218–239 (1911).

[2] Charzyński, Z., *Colloquium Math.*, **1,** 168–170 (1948).

[3] Löwner, K., *Math. Ann.*, **89,** 103–121 (1923).

[4] Pick, G., *Wien. Ber.*, **126,** 247–263 (1917).

[5] Schaeffer, A. C., Schiffer, M., and Spencer, D. C., *Duke Math. J.*, **16,** 493–527 (1949).

[6] Schaeffer, A. C., and Spencer, D. C., "Coefficient Regions for Schlicht Functions." (Forthcoming in Colloquium series.)

[7] Teichmüller, O., *Preuss. Akad. Wiss., Berlin, Sitzungsberichte*, **1938,** 363–375.

[8] The characteristics of $V_n$ were first discussed in reference (5).

PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES

Volume 35 December 15, 1949 Number 12

---

# AEROBIC FORMATION OF FUMARIC ACID IN THE MOLD RHIZOPUS NIGRICANS: SYNTHESIS BY DIRECT $C_2$ CONDENSATION*

BY J. W. FOSTER,† S. F. CARSON, D. S. ANTHONY, J. B. DAVIS,‡ W. E. JEFFERSON AND M. V. LONG

DEPARTMENT OF BACTERIOLOGY, UNIVERSITY OF TEXAS, AUSTIN, TEXAS, AND BIOLOGY DIVISION, OAK RIDGE NATIONAL LABORATORY, OAK RIDGE, TENNESSEE

Communicated by S. A. Waksman, October 21, 1949

Recent studies[1] have demonstrated that fumaric acid formation from glucose by *Rhizopus nigricans* No. 45 involves at least two mechanisms, one of which is aerobic, the other anaerobic. The latter involves a bulk fixation of $CO_2$ via oxalacetate, in confirmation of the reaction qualitatively demonstrated in this mold eight years ago with radioactive carbon dioxide ($C^{11}O_2$).[2]

The aerobic mechanism is the subject of the present work. Methods of cultivation and handling of the mold, submerged mycelium and analytical procedures are those given in detail by Foster and Davis[1, 3] and additional details will be given where necessary.

*Experiments and Results.—Relation of $C_2$ Compounds to Fumarate Formation from Glucose:* Using washed submerged mycelium the essential surface culture results of Butkewitsch and Federoff[4, 5] and Foster and Waksman[6] were confirmed, namely: aerobically ethanol accumulates in the early stages of the carbohydrate utilization, and gradually disappears, with a concomitant increase in fumarate, implying that alcohol is an intermediate between glucose and fumarate. Also confirmed was the formation of fumarate from alcohol as the sole carbon source, as well as from acetate, first noted by Takahashi and Asai in 1927.[7] A systematic study of fumarate formation from $C_2$ compounds (an aerobic process) was, therefore, undertaken.

*Conversion of Alcohol to Fumarate:* Extensive experimentation indicates the following conditions are essential for high efficiency of this conversion: (*a*) good aeration conditions, such as agitating thin layers on a reciprocal-type shaking machine; (*b*) initial alcohol concentration of 2 to 3%; (*c*) the growth medium in which the mycelium for the alcohol experiment is obtained must be neutralized to prevent a sharp drop in pH due to fumarate formation from the sugar; if the pH falls below 5.0 the mycelium is inferior for this purpose; (*d*) phosphate cannot be used for the neutralization because it yields inferior mycelium; excess precipitated calcium carbonate has proved very satisfactory.

The washed mycelium, which is now furnished alcohol, must be allowed to become acid. In the neutral range alcohol consumption is less, and conversion to fumarate is lacking altogether or is small. Enough $CaCO_3$ can be carried over from the growth medium, trapped in the mycelial clumps, or adhering to the mucilaginous coatings on the hyphae, so as to maintain the pH high enough to suppress fumarate formation from alcohol. Washing the mycelium briefly two or three times in 0.2 *N* HCl dissolves this $CaCO_3$ without injuring the mycelium. Excessive contact with the HCl is injurious. After a water wash a portion of the mycelium placed in water and agitated violently should show no $CaCO_3$ shaken loose and settling out. The optimum acidity for conversion of alcohol to fumaric acid is about pH 3 to 4, yet maximum conversions are obtained only when the pH initially is 6 to 7 and is allowed to fall due to fumaric acid accumulation in the now unbuffered medium; under good conditions the value falls to about pH 2.5 before the conversion of alcohol is affected harmfully.

In properly conducted experiments aseptic techniques are unnecessary, due to the high acidity which develops. In these experiments, 1.2 g. wet weight (= 0.2 g. dry wt.) mycelium was used per 25 ml. of 2% ethanol in 250-ml. Erlenmeyer flasks. After 4 to 5 days shaking at 28°C. and correction for small evaporation losses, filtrates from these cultures contain from 4 to 6 mg. free fumaric acid per ml. The highest we have obtained is 7.9 mg. per ml. Doubtless higher amounts could be obtained by partial neutralization. Generally ethanol consumption amounts to 4 to 10 mg. per ml. Analytical data show this is equivalent to about 50 to 80% weight conversion of ethanol to fumaric acid. This represents approximately 40 to 60% molar conversion efficiency. Actually, the three highest values we have observed have been weight conversion = 84, 86 and 91% (= molar conversions of 67, 68 and 72%). Fumarate is never obtained from mycelium in the absence of substrate (alcohol).

Two things are remarkable about this process—the high yields of fumaric acid and the purity of the product. Paper chromatography (solvent = butanol-propionic acid)[8] of a culture filtrate containing about 500 μg. fumaric acid showed no other detectable acid spots under our conditions. This means that other metabolic acidic impurities were present in amounts less than about 1% of the amount of fumaric acid. However, after substantial concentration, namely, ether extraction and crystallization of fumaric acid from a water solution of the residue, malic acid could be identified in the mother liquor from the crystallization. It showed an acid spot of RF = 0.43 on a paper chromatogram. The purified fumaric acid was identified as such by melting point and the $KMnO_4$ unsaturation test.

*Acetate Conversion:* The conversion of acetate to fumarate has been studied under a great variety of conditions, the best weight conversion never exceeding 30% (= 29% on molar basis). Neutral conditions are essential here, as utilization of acetate is greatly reduced on the acid side and fumarate formation is not observed at all. In confirmation of Butkewitsch and Federoff's[9, 10] studies with surface pads, some succinic acid is also formed with submerged mycelium. In one of our experiments 7.2 mg. acetate per ml. was consumed, yielding 1.92 mg. fumarate and 0.26 mg. succinate. The succinate was measured manometrically using a succinic dehydrogenase preparation from pig heart.

The discrepancy in yields from alcohol and acetate make it likely that acetic acid itself is not the substance undergoing conversion to fumarate, at least in the pathway from ethanol. An active $C_2$ compound probably is generated more efficiently from ethanol than from acetate, and though we have established that phosphate is essential for the conversion,[11] assiduous tests for acetyl phosphate[12] have been consistently negative, both on filtrates and ground mycelium extracts.

*Conversion of Succinate:* Fumarate formation from $C_2$ compounds and the fact that succinate accumulates in acetate cultures implies the intermediary formation of succinate followed by dehydrogenation to fumarate.

Demonstration of the dehydrogenation reaction is, therefore, essential for the premise that fumarate formation from $C_2$ compounds proceeds via succinate. Living *Rhizopus nigricans* mycelium is inert toward succinic acid, when tested manometrically for oxygen uptake. This is true at acid or neutral pH values. However, desiccated mycelium actively oxidizes succinate, indicating that permeability limitations account for the negative results with living mycelium. Judging from the amount of oxygen taken up, the oxidation of succinate does not proceed beyond fumarate in desiccated preparations, which by themselves display some endogenous $O_2$ consumption. In one experiment with 0.01 mM. succinate in phosphate buffer at pH 7.4, the oxygen uptake leveled off sharply to the endogenous rate, at which point the following $O_2$ uptake values were obtained: endogenous flask = 96 μl.; succinate flask = 200 μl.; difference due to succinate = 104 μl. $O_2$; theoretical for oxidation of succinate to fumarate = 112 μl. $O_2$. Each vessel had 30 mg. dry mold material.

*Experiments on $CO_2$ Fixation:* Because this organism can synthesize appreciable amounts of fumarate via $CO_2$ fixation with glucose as the substrate,[1] examination of $CO_2$ fixation mechanisms appeared desirable in the alcohol experiments. A double fixation, namely, $C_2 + CO_2 \rightarrow C_3$; $C_3 + CO_2 \rightarrow C_4$ would be involved, and of exceptional efficiency. The following experiments appear to preclude these mechanisms: (*a*) fumarate formation is undiminished in a $CO_2$-free atmosphere secured in a closed system by the presence of alkali and partial vacuum; (*b*) fumarate formation is not accelerated or enhanced by elevated $CO_2$ tensions; (*c*) experiments with radioactive carbon dioxide ($C^{14}O_2$) and unlabeled ethanol show that $CO_2$ could not possibly account for the major portion of the total fumarate formed. Acid-permanganate degradation (see later) of the fumaric acid formed in this experiment showed that the radioactivity was predominantly in the carboxyl groups as compared to methine carbon. In view of the fact that this mold has been demonstrated to possess oxalacetic acid decarboxylase, it is almost certain that the carboxyl radioactivity results from reversible decarboxylation of oxalacetate arising from fumarate via malate. Carbon dioxide is, therefore, a negligible factor in synthesis of fumarate from alcohol.

*Theoretical Consideration of the Above Results with Respect to a Tricarboxylic Acid Cycle Mechanism:* For two reasons the yield data seem to eliminate the tricarboxylic acid cycle in the formation of the $C_4$ dicarboxylic acids (fumaric).

First, theoretical molar yields of $C_4$ from $C_2$ via a $C_6$-cycle are 67%. In the mold system, yields equaling and possibly exceeding these have been obtained. That the Krebs cycle could operate in this case at its theoretical efficiency in producing $C_4$ is improbable, due to other degradation outlets for the intermediates. The fumarate yields from alcohol are in reality substantially higher than those reported above because they are based on total alcohol which had disappeared, after corrections from suitable evaporation controls. Actually some of the alcohol was assimilated by the mold mycelium. In one alcohol experiment the dry weight increase in mold mycelium was equivalent to 15% of the alcohol which had been consumed. Since this occurred in the absence of a nitrogen source and minerals, it is considered to represent oxidative assimilation.

Second, for each molecule of $C_4$ formed, the tricarboxylic acid cycle would require a $C_4$ molecule to begin with. This would have to come from ethanol, i.e., a $C_2$, which is the only substrate available. Thus, a functioning Krebs cycle in this system would still require a $C_4$ synthesis from $C_2$ by some extremely efficient mechanism.

Since $CO_2$ fixation has been eliminated (see above) as a major mechanism, attention was, therefore, directed to a consideration of a $2C_2$-condensation origin of the fumarate. None of the above experimental data are inconsistent with this possibility, and indeed, the yield data are strongly suggestive of this mechanism. Labeled ethanol makes it possible to put this theory to the final test.

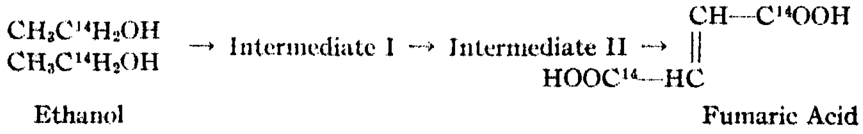
*Experiments Using $C^{14}$-Labeled Ethanol.—Methods:* Methyl-labeled ethanol and carbinol-labeled ethanol were obtained from Tracerlab, Inc. To avoid volatilization hazards, as well as to obtain $CO_2$ measurements, these experiments were conducted in 10-liter desiccators with ordinary air atmosphere. Other details were exactly as in the previous experiments with non-labeled ethanol.[1, 8] The cocks were opened daily to compensate for oxygen consumption. Only one flask was placed in a desiccator. The whole apparatus, with the flask fastened in position was placed on a shaker. Both kinds of labeled ethanol were run with portions of the same batch of mycelium and at the same time for 4 days. The radioactive alcohols were diluted with ordinary ethanol to a final concentration of 2%. At the end of the experiments a 300-ml. aliquot of the gas phase was removed with an evacuated bulb, the $CO_2$ absorbed in alkali, precipitated and weighed as $BaCO_3$ on a porcelain filter disc. Radioactivity measurements were made directly from the disc with a mica window Geiger-Müller counter. All radioactivity measurements on $BaCO_3$ were measured as close as possible to infinite thickness and corrected for self-absorption when necessary. All counts actually measured were at least 10 times the background, and most were 20 to 200 times. Radioactivity of other organic compounds was measured as $BaCO_3$ obtained from wet oxidation using persulfate-$AgNO_3$.[13] The liberated $CO_2$ was absorbed in NaOH and collected as $BaCO_3$ by addition of $BaCl_2$.

At time of analysis the culture with methyl-labeled ethanol had 5.4 mg. fumaric acid per ml., and that with carbinol-labeled ethanol 5.7 mg. per ml., determined by a mercurous fumarate method modified from Stotz.[14] Aliquots of the clear culture filtrates were adjusted to pH ~10 and about two-thirds the volume distilled off. The residue was acidified to pH ~2.0 with $H_2SO_4$ and extracted with ether overnight in a Kutscher-Steudel extractor. The ether in the extract was evaporated, and all remaining traces of volatile matter removed by maintaining the residue under high vacuum at 80°C. for 30 minutes. The light tan deposit of fumaric acid was dissolved in hot water, filtered and concentrated in an air stream on a hot plate to incipient crystallization. After standing in a refrigerator overnight, the crystals of radioactive fumaric acid were collected and washed on a sintered glass funnel. Yield from the carbinol-labeled alcohol was 52 mg. and from the methyl-labeled ethanol 32 mg. Both preparations melted at 274–275°C. in a sealed tube as compared to 279.5°C. for pure fumaric acid under our conditions.

About 5 mg. of each kind of labeled fumaric acid was combusted for total radioactivity and another 10 to 12 mg. degraded with acid-permanganate for location of the labeled carbons. Allen and Ruben[15] have shown that this oxidation liberates 3 moles $CO_2$ and 1 mole of formic acid per mole of fumarate, and that the formic acid is derived exclusively from one of the methine-carbons of the fumarate. The other methine-carbon and the two carboxyls make up the three $CO_2$ liberated. In confirmation, we have, with unlabeled fumarate, identified formic acid as a product of this oxidation under our conditions by ether extraction and its Duclaux distillation constants. Formate separation also has been done by steam distillation direct from the oxidation mixture after destruction of the residual $KMnO_4$ with $H_2O_2$. The oxidation mixture contained 11 mg. fumaric acid in 6 ml., 20 ml. 0.3 $N$ $KMnO_4$ and 14 ml. 3 $N$ $H_2SO_4$. The oxidation is complete in less than 10 minutes at room temperature, but the formate slowly is further oxidized. The oxidations were carried out in 160-ml. capacity Warburg vessels connected to manometers containing mercury, and when theoretical amounts of $CO_2$ had been liberated (= three-fourths of the fumarate carbon) the flask was placed in ice water to retard formate oxidation. Alkali was introduced through the vent in one side arm, and the $CO_2$ absorbed. The oxidation mixture was removed for formate separation, and the $CO_2$ precipitated as $BaCO_3$ and taken for radioactivity measurements.

In the case of radioactive fumarate, inactive carrier formic and acetic acids (0.3 mE. of each) were added prior to ether extraction and Duclaux distillation. Two successive extractions with 10 volumes of ether netted about 90% recovery of the acids. The ratios of acetic/formic acid obtained in the Duclaux distillations from both kinds of labeled fumaric acid were so shifted by the ether extracts that formic acid definitely must have been contributed by the degradation of fumaric acid. The Duclaux constants are indispensable in ascertaining the radioactivity of the formic acid produced from fumarate. Radioactivity of the various Duclaux fractions should be proportional to the characteristic formic acid values in a Duclaux distillation, otherwise the radioactivity in the preparation cannot be due to formic acid. After the titration values were obtained, the various Duclaux fractions were oxidized to $CO_2$ for radioactivity determinations.

*Results with Fumaric Acid from Carbinol-Labeled Ethanol:* The formic acid in this case was devoid of radioactivity even when measured as gaseous $CO_2$ with the vibrating reed electrometer which is $\sim$10 times more sensitive than our solid counting procedure. There was a small amount of steam distillable radioactivity ($\sim$2% of the total) but it doubtless came from impurities in the fumarate, for it bore no relation to the formate distillation curve. The fumaric acid itself had a specific activity of 1.2 (counts/sec./mg. $BaCO_3$), located entirely in the carboxyls, in view of the inactivity of methine-carbon. *This experiment demonstrates conclusively that a direct $C_2$ condensation has taken place, and furthermore, that no other reactions took place whereby carbinol-carbon of ethanol is converted to methine-carbon of fumaric acid.* That is, no mixing takes place. It means the following conversion must have occurred:



$$\begin{array}{ccc} \begin{matrix} CH_3C^{14}H_2OH \\ CH_3C^{14}H_2OH \end{matrix} & \rightarrow \text{Intermediate I} \rightarrow \text{Intermediate II} \rightarrow & \begin{matrix} CH—C^{14}OOH \\ \| \\ HOOC^{14}—HC \end{matrix} \\ \text{Ethanol} & & \text{Fumaric Acid} \end{array}$$

Based on the fact that phosphate and succinate appear to be involved, according to the earlier experiments, the supposition is made that Intermediate I is an "active" $C_2$ compound, possibly phosphorylated, and Intermediate II is the primary product of the condensation reaction, presumably an "active" succinate.

The starting radioactive alcohol used in this experiment had a specific activity of 1.67. The reduced specific activity (1.2) of the fumarate formed from this alcohol could result only from a dilution of the radioactive carboxyls by inactive carboxyls. The probable mechanism: some of the alcohol is oxidized to $CO_2$. The $CO_2$ at the end of the experiment had a specific activity of 1.2, representing inactive $CO_2$ from methyl carbons as well as endogenous respiration, and radioactive carbon from carbinol. Thus, due to dilution, the specific activity of the $CO_2$ is much less than that of the carbinol group of the alcohol ($2 \times 1.67 = 3.3$). After fumarate is formed via direct $C_2$ condensation, the initially highly radioactive carboxyl groups are diluted with lower specific activity $CO_2$, probably via reversible exchange by oxalacetate decarboxylase, an enzyme known to be active in this organism.[1]

*Results with Fumaric Acid from Methyl-Labeled Ethanol:* The formic acid derived from this fumarate was decidedly radioactive, as was the $CO_2$ representing three-fourths of the fumarate. This latter contains, of course, one methine carbon, equivalent to the formic acid carbon. The radioactivity was unquestionably due to formic acid, since radioactivity measurements precisely paralleled titration values of the Duclaux fractions representing pure formic acid. The last fraction, the most accurate because of its size, contained by titration 65% of the total formic acid distilled. This fraction also was found to contain 65% of the total radioactivity of the formic acid.

The specific activity of the starting ethanol in this experiment was 1.83, and that of the isolated fumaric acid 2.18. This increase can be accounted for only by entrance of labeled carbon into the initially non-radioactive carboxyls of the fumarate. Here, also, an exchange between labeled $CO_2$ and the fumarate carboxyls must have occurred, again probably involving the participation of the oxalacetate decarboxylase in this organism. The labeled $CO_2$ had a specific activity of 0.8, and was derived from oxidation of some ethanol. The incorporation of $C^{14}O_2$ into unlabeled carboxyl naturally results in increased specific activity of the fumarate. It is interesting that the exchange reaction in the carbinol-labeled alcohol experiment reduced the specific activity of the fumarate formed, and increased it in this experiment. These experimentally observed results are, therefore, what one would expect.

The specific activities of the methine-carbon and carboxyl-carbon of this fumarate were 3.6 and 0.7, respectively, as determined by measurements on formate and $CO_2$ from permanganate oxidation. Since the specific activity of the methyl group of the starting alcohol was determined to be 3.66, it is evident that a *direct conversion of the ethanol methyl group to fumarate methine group must have occurred.*

*Discussion.*—In recent years experiments on bacteria, yeast, fungi and animal tissue have suggested strongly the existence of the 2C condensation reaction (Thunberg-Wieland condensation). Although the evidence has, up to this point, been held in some dispute (see comprehensive literature reviews and critical analyses by Wood,[16] and Bloch,[17]) the experiments in this paper decisively establish the $2C_2$ condensation reaction. Thus, a third mechanism of formation of the $C_4$ dicarboxylic acids is added to the other two generally accepted as functioning in aerobic cellular respiration: $CO_2$ fixation yielding oxalacetic acid, and decarboxylation of α-ketoglutaric yielding succinate. Just what part, if any, the $2C_2$ condensation plays in various living systems in furnishing (eventually) oxalacetic acid for the tricarboxylic acid respiratory cycle is yet to be ascertained.

Apart from the reaction itself, of particular significance appear to be the results indicating entrance of $CO_2$ into carboxyl groups of the fumaric acid which itself is synthesized via a mechanism *not* concerned with $CO_2$ fixation. This $CO_2$ entrance is assumed to occur through reversible decarboxylation of pre-existing oxalacetate (or malate) in equilibrium with the fumarate. It would be of interest to determine if this mode of entrance of $CO_2$ could account for some of the carboxyl-labeled organic acids formed in the numerous Krebs type experiments which have heretofore been done with labeled $CO_2$ or $NaHCO_3$.

In view of the prominent rôle $C_2$ compounds play in the metabolism of fungi,[18] it is not surprising that the "$2C_2$ condensation reaction" should have a special significance in these organisms. With adequate oxygen, carbohydrate in mold cultures is oxidized via oxidative decarboxylation of pyruvate, the resulting active $C_2$ (acetyl phosphate?) forming organic acids and/or $CO_2$ eventually. However, the physical nature of mold mycelium in masses almost automatically limits oxygen availability during carbohydrate utilization, and anaerobic alcoholic fermentation ensues. Upon exhaustion of carbohydrate, the mold now slowly oxidizes the alcohol, generating organic acids and/or $CO_2$ just as it does from pyruvate directly. This mechanism whereby the mold eventually obtains its full quota of energy and carbon from the original carbohydrate and accumulates acids is possible because ethanol oxidation yields apparently the same or similar active $C_2$ fraction as pyruvate oxidation. The oxidation merely proceeds via a roundabout route, which has been named the "alcohol excursion." These results also indicate why alcohol has so often been considered disputably the intermediate in organic acid formation from sugars by fungi.

The formation of pure fumaric acid in high yields labeled predominantly in methine-carbons should make possible the availability of similarly labeled analogs, such as oxalacetic acid, malic acid, succinic acid, aspartic acid, α-ketoglutaric acid, etc., having value as tracers in metabolism studies. It also provides now the only feasible method of synthesizing methine-labeled maleic anhydride, a vital intermediate in chemical synthesis of many suitably labeled ring-type organic compounds, including carcinogens.

*Summary.*—Conditions for obtaining high yields of fumaric acid from ethanol by *Rhizopus nigricans* are specified. Yields of fumaric acid were so high that the tricarboxylic acid cycle as a mechanism of formation of fumarate is excluded. Yields of fumarate from acetate are one-third those from alcohol. Experiments with unlabeled ethanol and $C^{14}$-labeled carbon dioxide eliminate a $CO_2$ fixation mechanism of fumarate synthesis. The mold possesses an enzyme which oxidizes succinate to fumarate.

Fumarate formed from $C^{14}$ methyl-labeled ethanol contains the same specific activity in the methine-carbon as that of the methyl group from the starting ethanol. The carboxyls contain a small amount of radioactivity due to a reversible decarboxylation reaction probably involving oxalacetate decarboxylase.

Fumarate from $C^{14}$ carbinol-labeled ethanol contained no radioactivity in the methine-carbons, but abundantly in the carboxyls. All the evidence proves unequivocally that the $C_4$ dicarboxylic acid synthesis occurs by direct $2C_2$ condensation as hypothesized originally by Thunberg. The implications of this reaction in the tricarboxylic acid cycle of respiration and in synthesis of tracer intermediates for biochemistry and synthetic organic chemistry are discussed.

*Acknowledgment.*—One of the authors (J. W. Foster) wishes to express his gratitude to Doctor Alexander Hollaender, Director of the Biology Division, for the invitation which made it possible to come to Oak Ridge National Laboratory to participate in the tracer experiments.

---

* Supported in part by grants to J. W. Foster at the University of Texas by Ciba Pharmaceutical Laboratories, Inc., and the American Cancer Society, Inc. Work at the Oak Ridge National Laboratory performed under Contract No. W-7405-Eng-26 for the Atomic Energy Commission.

† On leave of absence from the University of Texas.

‡ Bacteriology Department, University of Texas.

[1] Foster, J. W., and Davis, J. B., "Anaerobic Fumarate Formation in the Mold *Rhizopus nigricans*," *J. Bact.*, 56, 329–338 (1948).

[2] Foster, J. W., Carson, S. F., Ruben, S., and Kamen, M. D., "Radioactive Carbon as an Indicator of Carbon Dioxide Utilization. VII. The Assimilation of Carbon Dioxide by Molds," these PROCEEDINGS, **27,** 590–596 (1941).

[3] Foster, J. W., and Davis, J. B., "Carbon Dioxide Inhibition of Anaerobic Fumarate Formation in the Mold *Rhizopus nigricans*," *Arch. Biochem.*, **21,** 135–142 (1949).

[4] Butkewitsch, W., and Federoff, M., "Über Bildung von Fumarsäure in den Zuckerkulturen von *Mucor stolonifer* (*Rhizopus nigricans*) und sein Verhalten zur Brenztraubsäure," *Biochem. Zt.*, **206,** 440–456 (1929).

[5] Butkewitsch, W., and Federoff, M., "Über die Umwandlung der Äthylalkohole in den Kulturen von *Mucor stolonifer*," *Ibid.*, **219,** 103–121 (1930).

[6] Foster, J. W., and Waksman, S. A., "The Specific Effect of Zinc and Other Heavy Metals on Growth and Fumaric Acid Production by *Rhizopus*," *J. Bact.*, **37,** 599–617 (1939).

[7] Takahashi, T., and Asai, T., "Acids Formed by Species of *Rhizopus*," *Proc. Imp. Acad. Sci., Japan*, **3,** 86–89 (1927).

[8] Benson, A. A., Bassham, S. H., Calvin, M., Goodale, T. C., Haas, V. A., and Stepka, N., "The Pathway of Carbon in Photosynthesis. Paper Chromatography and Radioautography of the Products." Abstracts of Papers, Am. Chem. Soc. Meetings, Atlantic City, pp. 58C–59C, 1949.

[9] Butkewitsch, W., and Federoff, M., "Über die Umwandlung der Essigsäure durch *Mucor stolonifer* in Bernstein- und Fumarsäure und das Verfahren zur Trennung und quantitative Bestimmung dieser Säuren," *Biochem. Zt.*, **207,** 302–318 (1929).

[10] Butkewitsch, W. and Federoff, M., "Über die Verhaltnisse zwischen Essig-, Bernstein-, Fumar- und Oxalsäure in den Kulturen von *Mucor stolonifer* und einigen anderen Pilzen," *Ibid.*, **219,** 87–102 (1930).

[11] Foster, J. W., and Davis, J. B., Unpublished data.

[12] Lipmann, F., and Tuttle, L. C., "A Specific Micromethod for the Determination of Acyl Phosphates," *J. Biol. Chem.*, **159,** 21–28 (1945).

[13] Osburn, O. L., and Werkman, C. H., "Determination of Carbon in Fermented Liquors," *Ind. Eng. Chem., Anal. Ed.*, **4,** 421–423 (1932).

[14] Stotz, E., "Determination of Fumaric Acid in Protein Solutions Containing Succinic Acid," *J. Biol. Chem.*, **118,** 471–477 (1937).

[15] Allen, M. B., and Ruben, S., "Tracer Studies with Radioactive Carbon and Hydrogen. The Synthesis and Oxidation of Fumaric Acid," *J. Am. Chem. Soc.*, **68,** 948–950 (1942).

[16] Wood, H. G., "The Fixation of Carbon Dioxide and the Interrelationships of the Tricarboxylic Acid Cycle," *Physiol. Rev.*, **26,** 198–246 (1946).

[17] Bloch, K., "Metabolism of Acetic Acid in Animal Tissues," *Ibid.*, **27,** 574–620 (1947).

[18] Foster, J. W., *Chemical Activities of Fungi*, Academic Press, Inc., New York, 1949.

# GEOMETRY OF REDUCED MOMENT SPACES

By S. Karlin and L. S. Shapley

The RAND Corporation, Santa Monica, California

Communicated by J. von Neumann, October 7, 1949

1. A normalized distribution function on the closed unit interval [0, 1] is any real, non-decreasing function $\phi(t)$, continuous to the right, with $\phi(-0) = 0$, $\phi(1) = 1$. The distribution with saltus of one at $t$ and constant elsewhere is denoted by $I_t$. The $n$th *moment* of $\phi$ is

$$\mu_n(\phi) = \int_{-0}^{1} t^n d\phi(t).$$

The region of $n$-dimensional Euclidean space comprising the points $x_i = \mu_i(\phi)$, $i = 1, 2, \ldots, n$, for all distributions $\phi$ is the *reduced moment space* $\mathfrak{D}^n$. The region comprising the points of the form $x_i = \mu_i(I_t) = t^i$ is a twisted curve, denoted by $\mathfrak{C}^n$. $\mathfrak{D}^n$ is closed and convex, and for $n \geq 2$ its set of extreme points is precisely $\mathfrak{C}^n$.

Theorem. *If $x$ is in the boundary of $\mathfrak{D}^n$ then there is a unique distribution $\phi$ with moments $\mu_i(\phi) = x_i$, and $x$ can be represented as a convex combination of extreme points in just one way, namely*

$$x_i = \int_{-0}^{1} \mu_i(I_t) d\phi(t).$$

*If $x$ is interior to $\mathfrak{D}^n$, then the associated distribution and the convex representation are not unique.*

Denote by $\bar{b}(x)$ the number of extreme points participating in the convex representation of a boundary point $x$ of $\mathfrak{D}^n$, and $b(x)$ the same but with the points $(0, 0, \ldots, 0)$ and $(1, 1, \ldots, 1)$ counted half. Let $\mathfrak{E}(x)$ be the common intersection of all the supporting planes to $\mathfrak{D}^n$ at $x$; let $a(x)$ denote the dimension of $\mathfrak{E}(x)$, $c(x)$ the dimension of $\mathfrak{E}(x) \cap \mathfrak{D}^n$.

Theorem. *If $x$ is any boundary point of $\mathfrak{D}^n$ then*

$$2b(x) = a(x) + 1,$$
$$\bar{b}(x) = c(x) + 1.$$

By a suitable extension of the definitions, these relations can be made to hold for points interior and exterior to $\mathfrak{D}^n$. In particular, there are two distinct "minimal" representations of the interior points, with $b(x) = (n + 1)/2$, corresponding to the representations of two boundary points of $\mathfrak{D}^{n+1}$. The boundary of $\mathfrak{D}^n$ may be partitioned into $2n$ disjoint "faces," $\overline{\mathfrak{A}}_a^n$ and $\underline{\mathfrak{A}}_a^n$, of surface dimension $a$, $a = 0, 1, \ldots, n - 1$, where $a(x) = a$ for $x \in \overline{\mathfrak{A}}_a^n$ or $\underline{\mathfrak{A}}_a^n$. $c(x)$ is constant over these faces, and indicates the order of curvature of the surface. For the entire boundary, $a - 1 \leq 2c \leq a + 1$.

2. The set of polynomials $P(t)$ of degree $n$ or less which are non-negative over $[0, 1]$ may be conceived as a conical region in the $(n + 1)$-dimensional Euclidean space of the coefficients. If the polynomials are normalized by

$$\int_0^1 P(t)dt = 1$$

a convex, closed, $n$-dimensional region $\mathfrak{P}^n$ is obtained. Its extreme points are precisely those polynomials that have $n$ roots (counting multiplicity) in $[0, 1]$: they fall into two disjoint components distinguished by the sign of the leading coefficient. The entire region $\mathfrak{P}^n$ is spanned by extreme points taken two at a time; in fact:

THEOREM. *If $P$ is in $\mathfrak{P}^n$ then there is a unique representation*

$$P(t) = \alpha \prod_{j=1}^{m} (t - t_{2j-1})^2 + \beta t(1 - t) \prod_{j=1}^{m-1} (t - t_{2j})^2, \textit{ if } n = 2m,$$

$$P(t) = \alpha t \prod_{j=1}^{m} (t - t_{2j})^2 + \beta(1 - t) \prod_{j=1}^{m} (t - t_{2j-1})^2, \textit{ if } n = 2m + 1,$$

*with $\alpha \geq 0$, $\beta \geq 0$, $0 \leq t_1 \leq \ldots \leq t_{n-1} \leq 1$. $P$ is interior to $\mathfrak{P}^n$ if and only if all of the inequalities are strict*

Denote by $b(P)$ the number of distinct roots of $P(t)$; $b(P)$ the same counting roots at 0 and 1 as half. Then for $P$ on the boundary of $\mathfrak{P}^n$:

$$\begin{aligned} \bar{b}(P) &= n - a(P) \\ 2b(P) &= n - c(P), \end{aligned}$$

$a(P)$ and $c(P)$ being defined geometrically as in section 1. The boundary of $\mathfrak{P}^n$ may be partitioned into $2n$ disjoint components $\overline{\mathfrak{C}}_c^n$ and $\underline{\mathfrak{C}}_c^n$, $c = 0, 1, \ldots, n - 1$, with $c(P) = c$ for $P$ in $\overline{\mathfrak{C}}_c^n$ or $\underline{\mathfrak{C}}_c^n$.

3. A *convex cone* in $(n + 1)$-dimensional space is a set which contains the point $\alpha x + \beta x'$ whenever it contains $x$ and $x'$, $\alpha$ and $\beta$ being any non-negative real numbers.[1] The *conjugate* cone $C^*$ to a convex cone $C$ is the set of points $y$ which satisfy

$$\sum_{i=0}^{n} x_i y_i \geq 0 \qquad \text{all } x \in C.$$

$C^*$ is convex closed, and if $C$ is closed then $C^{**} = C$.

If $\mathfrak{D}^n$ and $\mathfrak{P}^n$ are made into convex cones in $(n + 1)$-dimensional space by dropping the two normalizing conditions $\int d\phi = 1$, and $\int P dt = 1$, then they are mutually conjugate. Geometrically stated, this means that the coördinates of a boundary point of $\mathfrak{D}^n$ are the coefficients of a plane of support to $\mathfrak{P}^n$, and vice versa. Interior points correspond to separating planes, exterior points to secant planes. If $x$ is in the boundary of $\mathfrak{M}^n$ and if $P$ is interior to the (convex) set of points in the boundary of $\mathfrak{P}^n$ which generate planes of support to $\mathfrak{M}^n$ at $x$, then $P$ is said to be *conjugate* to $x$.

THEOREM. *If $x$ is conjugate to $P$, or $P$ conjugate to $x$, then*

$$\begin{aligned} a(x) + c(P) &= c(x) + a(P) = n - 1, \\ \underline{b}(x) = \underline{b}(P), &\qquad \bar{b}(x) = \bar{b}(P). \end{aligned}$$

The duality of conjugate cones is fundamental in the geometrical interpretation of the theory of two-person, zero-sum games. Motivation for the present study came from the consideration of games described by polynomial payoff functions over the unit square.

4. In this section, results will be given only for the case $n = 2m$. Analogous results hold for odd dimensions.

The duality just described leads easily to the well-known fact,[2] that $\mu_0 = 1, \mu_1, \ldots, \mu_{2m}$ are the moments of a distribution if and only if the quadratic forms

$$\sum_{i,j=0}^{m} x_i x_j \mu_{i+j} \text{ and } \sum_{i,j=1}^{m} x_i x_j (\mu_{i+j-1} - \mu_{i+j})$$

are non-negative. The first principal minors are

$$\underline{\Delta}_k = \begin{vmatrix} \mu_0 & \cdots & \mu_k \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ u_k & \cdots & \mu_{2k} \end{vmatrix} \text{ and } \bar{\Delta}_l = \begin{vmatrix} \mu_1 - \mu_2 & \cdots & \mu_l - \mu_{l+1} \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ \mu_l - \mu_{l+1} & \cdots & \mu_{2l-1} - \mu_{2l} \end{vmatrix}$$

$k = 0, 1, \ldots, m$, $l = 1, 2, \ldots, m$, respectively. These determinants may be used to characterize the faces $\overline{\mathfrak{A}}_a^{2m}$, $\underline{\mathfrak{A}}_a^{2m}$ of $\mathfrak{M}^{2m}$. In fact, if $x$ is in $\overline{\mathfrak{A}}_a^{2m}$ then, setting $b = (a+1)/2$,

$$\begin{aligned} &\underline{\Delta}_k > 0, \quad 1 \leq k \leq b; \qquad \underline{\Delta}_k = 0, \quad b < k \leq m; \\ &\bar{\Delta}_l > 0, \quad 1 \leq l < b; \qquad \bar{\Delta}_l = 0, \quad b \leq l \leq m. \end{aligned}$$

If $x$ is in $\underline{\mathfrak{A}}_a^{2m}$ then interchange upper and lower bars.

If an interior point $(\mu_1, \ldots, \mu_{2m-1})$ of $\mathfrak{M}^{2m-1}$ is given, then the extreme values possible for $\mu_{2m}$ are found by solving $\underline{\Delta}_{2m} = 0$ and $\bar{\Delta}_{2m} = 0$. The lower value, designated $\underline{\mu}_{2m}$ comes from the first equation, the upper, $\bar{\mu}_{2m}$, from the second. The supporting planes at the opposite boundary points $(\mu_0, \ldots, \underline{\mu}_{2m})$ and $(\mu_0, \ldots, \bar{\mu}_{2m})$ are unique and correspond to the polynomials $[P_{\underline{\Delta}_m}(t)]^2$ and $t(1-t)[P_{\bar{\Delta}_m}(t)]^2$ of $\mathfrak{P}^{2m}$ (when normalized), where

$$P_{\underline{\Delta}_m}(t) = \begin{vmatrix} \mu_0 & \dots & \mu_{m-1} & 1 \\ \mu_1 & & & t \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \mu_m & \dots & \mu_{2m-1} & t^m \end{vmatrix},$$

$$P_{\bar{\Delta}_m}(t) = \begin{vmatrix} \mu_1 - \mu_2 & \dots & \mu_{m-1} - \mu_m & 1 \\ \mu_2 - \mu_3 & & & t \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \mu_m - \mu_{m+1} & \dots & \mu_{2m-2} - \mu_{2m-1} & t^{m-1} \end{vmatrix}.$$

The spectra of the unique distribution $\underline{\psi}$ and $\bar{\psi}$ corresponding to these opposite boundary points are the roots of $P_{\underline{\Delta}_m}(t)$ and $t(1 - t)P_{\bar{\Delta}_m}(t)$, respectively. The two sets of roots interlock strictly (cf. the representation of section 2) and the cumulative weights of the two distributions interlock as well, so that $\underline{\psi}(t) - \bar{\psi}(t)$ has a sign change at each saltus point of $\underline{\psi}$ and $\bar{\psi}$.

The polynomials $P_{\underline{\Delta}_k}$, $k = 0, 1, \dots, m$, are orthogonal with respect to the weight factor $d\underline{\psi}(t)$, while $P_{\bar{\Delta}_l}$, $l = 1, 2, \dots, m$, are orthogonal with respect to $t(1 - t)d\bar{\psi}(t)$. These systems of polynomials, together with the analogs for odd $n$, are the basis for a new geometric approach to the theory of orthogonal polynomials.

THEOREM. *There is a 1:1:1 correspondence between* (*a*) *the interior of* $\mathfrak{D}^{2m-1}$, (*b*) *the set of ordered pairs of polynomials* $(P_{\underline{\Delta}_m}, P_{\bar{\Delta}_m})$, *and* (*c*) *the open simplex of strictly interlocking roots* $0 = t_0 < t_1 < \dots < t_{2m} = 1$.

The representation of section 2 is thus linked to the polynomials of this section and to the interior of $\mathfrak{D}^{n-1}$.

5. In this section certain geometric features of the reduced moment space $\mathfrak{D}^n$ and its dual $\mathfrak{P}^n$ will be given. $\mathfrak{D}^n$ can be inscribed in a very natural simplex $\mathfrak{S}^n$ whose vertices $x^{(k)}$ are given by

$$x_i^{(k)} = \frac{\binom{k}{i}}{\binom{n}{i}} \qquad i = 1, 2, \dots, n; \ k = 0, 1, \dots, n.$$

Its dual is the simplex inscribed in $\mathfrak{P}^n$ spanned by the polynomials

$$\lambda_{n,k}(t) = \binom{n}{k} t^k (1 - t)^{n-k} \qquad k = 0, 1, \dots, n,$$

(which must be normalized by the factor $n + 1$). The barycentric coordinates with respect to $\mathfrak{S}^n$ of a point $x$ of $\mathfrak{D}^n$ are

$$\lambda_{n,k} = \int_{-0}^{1} \lambda_{n,k}(t) d\phi(t) \qquad k = 0, 1, \ldots, n,$$

where $\phi$ is a distribution having moments $\mu_i(\phi) = x_i$, $i = 1, 2, \ldots, n$. This gives a geometric interpretation to the quantities $\lambda_{n,k}$ which play so important a rôle in the literature on moment theory.[3] The relation of $\mathfrak{S}^n$ to $\mathfrak{D}^n$ leads to several qualitative results on moment spaces, and in particular gives a simple geometrical proof of Hausdorff's Theorem.

An interesting point in the interior of each $\mathfrak{D}^n$ is that given by the continuous distribution $dt/\sqrt{t(1 - t)}$ whose moments

$$\mu_i = \frac{1 \cdot 3 \cdot \ldots \cdot 2n - 1}{2 \cdot 4 \cdot \ldots \cdot 2n}$$

satisfy, for any $n$, $\mu_n = (\underline{\mu}_n + \bar{\mu}_n)/2$. The supporting planes at $(\mu_1, \mu_2, \ldots, \underline{\mu}_n)$ and $(\mu_1, \mu_2, \ldots, \bar{\mu}_n)$ are parallel, another unique property of this moment sequence. The polynomials $P_{\underline{\Delta}_m}$, $P_{\bar{\Delta}_m}$, $n = 2m$, for these moments are (up to a constant factor) the Tchebycheff polynomials of the first and second kind.

The center of gravity of the simplex $\mathfrak{S}^n$ is given by the rectangular distribution $\phi(t) = t$, whose moments $1, {}^1/_2, {}^1/_3, \ldots$ give rise in similar fashion to the Legendre and the first associated Legendre polynomials. The centroid of the dual simplex is $P(t) \equiv 1$.

Finally, the $n$-dimensional volume of $\mathfrak{D}^n$ has the interesting value

$$V_n = \prod_{i=1}^{n} B(i, i),$$

where $B$ is the beta function.

6. A similar account can be made for moment spaces based on an arbitrary finite set of moments $\{\mu_{i_\nu}\}$. With certain adaptations, most of the foregoing analysis and geometry can be carried over to infinite moment sequences, and to distributions over the infinite intervals $[0, \infty]$ and $[-\infty, \infty]$. These results and their proof, and additional theorems unstated here, will be published elsewhere.

---

[1] Krein, M., and Šmulian, D., *Ann. Math.*, **41**, 556–583 (1940).

[2] Shohat, J. A., and Tamarkin, J. D., *The Problem of Moments*, American Mathematical Society, New York, 1943, p. 77.

[3] Widder, D. V., *The Laplace Transform*, Princeton University Press, Princeton, 1941, ch. III.

# ON $n$-DIMENSIONAL CONCEPTS OF BOUNDED VARIATION, ABSOLUTE CONTINUITY AND GENERALIZED JACOBIAN

By T. Rado and P. V. Reichelderfer

Ohio State University, Columbus

Communicated by Marston Morse, October 17, 1949

*Background.*—A real-valued function $x = f(u)$ may be thought of as defining a mapping $f\colon U^1 \to X^1$, where $U^1$, $X^1$ designate Euclidean 1-spaces (real number-lines). In this interpretation the fundamental concepts of bounded variation, absolute continuity and derivative acquire important and suggestive geometrical meaning. To obtain fruitful $n$-dimensional generalizations of these fundamental concepts, it is then natural to study mappings $f\colon U^n \to X^n$, where $U^n$, $X^n$ are Euclidean $n$-spaces with points $u = (u_1, \ldots, u_n), x = (x_1, \ldots, x_n)$. The general objective is to develop concepts of bounded variation, absolute continuity and generalized Jacobian, which yield a theory comparable in utility and scope with the classical one-dimensional theory. Various approaches to this general objective may be found in the literature. The present note is concerned with the line of thought presented in reference 1. The theory in reference 1 is developed for the case $n = 2$. For this case, the theory may be considered as essentially complete, due to the sustained efforts of many mathematicians. The generalization to the $n$-dimensional case represents a long-range program of great interest and great difficulty. The purpose of this note is to report on progress achieved by the writers in this general direction. We shall present in detail the precise definitions of the $n$-dimensional concepts of bounded variation, absolute continuity and generalized Jacobian which we use in our work, and we shall describe applications to the transformation of multiple integrals, a topic which is discussed in Part IV of reference 1 for the case $n = 2$.

*Basic Concepts.*—Let $D_0$ be a bounded domain (connected open set) in the space $U^n$, and let $f\colon D_0 \to X^n$ be a bounded continuous mapping from $D_0$ into the space $X^n$. In analogy with the 2-dimensional case it is expedient to assign to every point $x$ of $X^n$ an essential multiplicity relative to the mapping $f$. For the 2-dimensional case, various equivalent definitions of the essential multiplicity are available—in fact, four such essential multiplicities were actually used in the literature. In reference 2 these essential multiplicities are denoted by $\psi$, $\psi^*$, $\Psi$, $\Psi^*$, respectively. Since they differ from each other only on countable sets, they can be used interchangeably throughout the 2-dimensional theory. However, the corresponding essential multiplicities in $n$ dimensions may differ from each other on sets of positive measure. Recent results of Federer[3] indicate that the multiplicity corresponding to the largest one of the four multiplicities just mentioned, namely $\Psi^*$, is the appropriate one for the $n$-dimensional case. In the work of Federer, the definition of this multiplicity is stated, in a very convenient manner, in terms of relative Cech cohomology groups. This essential multiplicity is used in the present note, and is denoted by $K$. For the concept of absolute continuity several equivalent definitions are available in the 2-dimensional case. For the $n$-dimensional case the choice is again restricted; we found it appropriate to adopt the definition proposed in reference 4. Finally, for the $n$-dimensional generalized Jacobian we found it convenient to adopt the form of the definition suggested in reference 5 for the 2-dimensional case. Precise definitions of these basic concepts follow.

*The Index* $\mu(x, f, D)$.—Given a bounded continuous mapping $T: D_0 \rightarrow X^n$ as above, let $x_0$ be any point in $X^n$. A domain $D$ is termed admissible for $x_0, f, D_0$ if the closure $cD$ of $D$ is contained in $D_0$ and $x_0$ is not in the image of the frontier $frD$ of $D$ under $f$. If $D$ is admissible for $x_0$, $f$, $D_0$, there exist in $X^n$ pairs of compact sets $A$, $B$ satisfying the following conditions: (i) $frA$ is a subset of $B$, and $B$ is a subset of $A$; (ii) $A - B$ is not empty and is connected; (iii) the image of $cD$ under $f$ is contained in $A$; (iv) the image of $frD$ under $f$ is contained in $B$; (v) $x_0$ is in $A - B$. Under these conditions the pair $A$, $B$ is termed admissible for $x_0$, $f$, $D$. Assume that $A$, $B$ is admissible for $x_0$, $f$, $D$, and consider the mapping $f|cD: (cD, frD) \rightarrow (A, B)$. This mapping induces a homomorphism $h: H^n(A, B) \rightarrow H^n(cD, frD)$, where $H^n$ denotes the $n$-dimensional Cech cohomology group with integral coefficients. Both of the groups involved are infinite cyclic, and by a suitable orientation process one may select once for all standard generators $g_D$, $g_{AB}$ for the groups $H^n(cD, frD)$, $H^n(A, B)$, respectively. Then $h(g_{AB}) = k \cdot g_D$, where $k$ is an integer (positive, negative or zero). This integer $k$ is defined to be the index $\mu(x_0, f, D)$; it can be easily shown to depend solely upon $x_0$, $f$, $D$ as the notation suggests. If $\mu(x_0, f, D)$ is not zero, then $D$ is termed an indicator domain for $x_0, f, D_0$. An indicator domain $D$ is termed positive (negative) if the sign of $\mu(x_0, f, D)$ is positive (negative).

*The Essential Multiplicity Function* $K(x, f, D_0)$.—Let $\sigma$ be a generic notation for a finite system of pairwise disjoint indicator domains $D$ for $x_0$, $f$, $D_0$. Then $K(x_0, f, D_0)$ is defined to be the least upper bound of $\sum |\mu(x_0, f, D)|$ for $D$ in $\sigma$, with respect to all systems $\sigma$. If each $D$ occurring in the system $\sigma$ is restricted to be a positive (negative) indicator domain, one obtains the definition of the multiplicity functions $K^+(x_0, f, D_0)$, $(K^-(x_0, f, D_0))$.

*The Essential Maximal Model Continuum.*—A continuum $c$ is termed an essential maximal model continuum for $x_0$, $f$, $D_0$ if $c$ is a component of the inverse of $x_0$ under $f$ in $D_0$, and every neighborhood of $c$ contains in its interior an indicator domain $D$ which contains $c$. The essential set

$E(f, D_0)$ consists of all points $u_0$ in $D_0$ which belong to an essential maximal model continuum for $f(u_0)$, $f$, $D_0$.

*Essential Bounded Variation, Essential Absolute Continuity.*—The mapping $f$ is said to be *eBV* in $D_0$ (essentially of bounded variation in $D_0$) if $K(x, f, D_0)$ is summable. The mapping $f$ is said to be *eAC* in $D_0$ (essentially absolutely continuous in $D_0$) if (i) it is *eBV* in $D_0$, and (ii) for every subset $b$ of $E(f, D_0)$ which has measure zero, it is true that the measure of the image of $b$ under $f$ is also zero.

*The Signed Essential Multiplicity Function* $\mu_e(x, f, D_0)$. This is defined to be $K^+(x, f, D_0) - K^-(x, f, D_0)$ unless both $K^+$ and $K^-$ are infinite. When $f$ is *eBV* in $D_0$, $\mu_e(x, f, D_0)$ is defined almost everywhere in $X^n$ and is summable. If $D$ is an admissible domain for $x_0$, $f$, $D_0$, and $K(x_0, f, D)$ is finite, then the index $\mu(x_0, f, D)$ and the signed essential multiplicity $\mu_e(x_0, f, D)$ have the same value.

*The Essential Generalized Jacobian* $J_e(u, T)$.—Assume that $f$ is *eBV* in $D_0$. For each open interval $I$ in $D_0$, the signed essential multiplicity function $\mu_e(x, f, I)$ is defined almost everywhere in $X^n$ and is summable. The Lebesgue integral of $\mu_e(x, f, I)$ yields a function of intervals $I$ in $D_0$ which possesses a derivative almost everywhere in $D_0$; this derivative is the essential generalized Jacobian $J_e(u, f)$.

*Application to the Transformation of n-tuple Integrals.* In terms of the preceding concepts, transformation formulae for $n$-tuple integrals are established which represent complete generalizations of the 2-dimensional formulae in reference 1, IV.4.4, IV.4.6, IV.4.7, IV.4.9; with the signed essential multiplicity $\mu_e$ replacing the function $\nu$ in these formulae. As a generalization of reference 1, IV.4.8, it is shown that if $f$ is *eBV* in $D_0$, then the integral over $D_0$ of $|J_e(u, f)|$ is less than or equal to the integral over $X^n$ of $K(x, f, D_0)$, the sign of equality holding if and only if $f$ is *eAC* in $D_0$. The range of applicability of these $n$-dimensional transformation formulae depends upon the scope of the *eAC* mappings. It is shown that this class is closed under various limit processes, the results representing complete generalizations of the closure theorems in reference 1, IV.4.11, IV.4.12, IV.4.15, IV.4.16, IV.4.17, IV.4.18, IV.4.19. It is demonstrated that if $f$ is *eBV* in $D_0$ and if the coordinate functions defining $f$ possess complete differentials almost everywhere in $D_0$, then $J_e(u, f)$ agrees with the ordinary Jacobian almost everywhere in $D_0$. Consequently the $n$-dimensional transformation formulae contain, as very special cases, various special results established in previous literature. Many further results presented in Part IV of reference 1 admit of a complete generalization to $n$ dimensions.

A detailed presentation of our work is now being prepared for publication.

---

[1] Rado, T., "Length and Area," *American Mathematical Society Colloquim Publications*, vol. 30, New York, 1948.

[2] Rado, T., "Two-Dimensional Concepts of Bounded Variation and Absolute Continuity," *Duke Math. Jl.*, **14**, 587–608 (1947).

[3] Federer, H., "Essential Multiplicity and Lebesgue Area," Proc. Natl. Acad. Sci., **34**, 611–616 (1948).

[4] Rado, T., "On Essentially Absolutely Continuous Plane Transformations," *Bull. Am. Math. Soc.*, **55**, 629–632 (1949).

[5] Reichelderfer, P. V., "Law of Transformation for Essential Generalized Jacobians," *Duke Math. Jl.*, **16**, 73–83 (1949).

---

# *QUADRATIC RELATIONS INVOLVING THE NUMBERS OF SOLUTIONS OF CERTAIN TYPES OF EQUATIONS IN A FINITE FIELD*

By H. S. Vandiver

Department of Applied Mathematics, University of Texas

Communicated October 7, 1949

In two previous papers[1] the writer considered the problem of finding the number of sets $s$, $t$, such that

$$g^{i + m_1 s} + g^{j + m_2 t} + 1 = 0, \tag{1}$$

where $m_1$ and $m_2$ are integers $0 < m_1 < p^n - 1$; $0 < m_2 < p^n - 1$; $p^n - 1 \equiv 0 \pmod{m_1}$, $p^n - 1 \equiv 0 \pmod{m_2}$, $g$ is a primitive root in a finite field $F(p^n)$, $p$ an odd prime, $i$ a given integer in the set $0, 1, \ldots, m_1 - 1$; $j$ a given integer in the set $0, 1, \ldots, m_2 - 1$, $s$ in the set $0, 1, \ldots, m_1' - 1$; $t$ in the set $0, 1, \ldots, m_2' - 1$; $p^n - 1 = m_1m_1' = m_2m_2'$. Denoting the number of possible sets $s$, $t$ by $(i, j)_{m_1m_2}$, an explicit expression in terms of $p$ and $n$ for the expression

$$\sum_{\substack{i = 0, \ldots m_1 - 1 \\ j = 0, \ldots m_2 - 1}} (i, j)_{m_1m_2}(i + h, j + k)_{m_1m_2},$$

in the two Cases I. $m_1 = m_2$, II. $m_1m_2 = p^n - 1$, $(m_1, m_2) = 1$, was set up. In the present paper we obtain, using much simpler methods, similar results for $m_1$ and $m_2$ any divisors of $p^n - 1$, also quadratic relations of a different type when $m_1 = m_2$. In Hua and Vandiver[2] an expression for the number of solutions of (1) was derived, noting that the equation

$$c_1x_1^{m_1} + c_2x_2^{m_2} + c_3 = 0 \tag{2}$$

in $F(p^n)$ may be put in the form (1), when $c_1c_2c_3x_1x_2 \neq 0$, but the use of this form[2] is inconvenient for our present purpose. We have recourse to some results in another paper by the writer.[3]

If we keep $i$ and $j$ fixed the number of solutions of (1) is $m_1m_2(i, j)$ where $(i, j)$ is the number of different sets $s$, $t$ which are possible in the relation

(3) and in the respective ranges just mentioned for $s$ and $t$. Let now $(k, l)_m$ be the number of different solutions $s_1$, $t_1$ of

$$1 + g^{k + ms_1} + g^{l + mt_1} = 0, \tag{3}$$

with $m$ the L. C. M. of $m_1$ and $m_2$, with both $s_1$ and $t_1$ in the range 0, 1, ... $((p^n - 1)/m) - 1$. Consider the expression

$$\theta(\alpha) = \sum_{k,\, l}^{0 \text{ to } m - 1} (k, l)_m \alpha^M, \tag{4}$$

where $\alpha = e^{2i\pi/m}$, and $M = u_1km/m_1 + u_2lm/m_2$. This may also be written as

$$\sum_{k,\, l}^{0 \text{ to } m - 1} (k, l)_m \alpha_1^{u_1k} \alpha_2^{u_2l},$$

where $\alpha_1 = e^{2i\pi/m_1}$, $\alpha_2 = e^{2i\pi/m_2}$. Further this may be reduced to[3]

$$\psi(\alpha_1^{u_1}, \alpha_2^{u_2}) = \sum_{i,\, j} (i, j) \alpha_1^{u_1i} \alpha_2^{u_2j}, \tag{5}$$

where $i$ ranges independently over $0, 1, \ldots, m_1 - 1$ and $j$ independently over $0, 1, \ldots, m_2 - 1$, with $(k, l) = (k, l)_{m_1m_2}$ and

$$\psi(\alpha_1^{u_1}, \alpha_2^{u_2}) \psi(\alpha_1^{-u_1}, \alpha_2^{-u_2}) = p^n, \tag{6}$$

with $\alpha_1^{u_1} \neq 1$, $\alpha_2^{u_2} \neq 1$, $\alpha_1^{u_1}\alpha_2^{u_2} \neq 1$, since it is known that $\theta(\alpha)\theta(\alpha^{-1}) = p^n$ under the conditions, $u_1\frac{m}{m_1} \not\equiv 0 \pmod{m}$; $u_2\frac{m}{m_2} \not\equiv 0 \pmod{m}$; and $u_1\frac{m}{m_1} + u_2\frac{m}{m_2} \not\equiv 0 \pmod{m}$, which reduce to the former conditions. Now for a fixed $i$ and $j$, with $i'$ and $u_1$ ranging independently over $0, 1, \ldots, m_1 - 1$, and $j'$ and $u_2$ ranging independently over $0, 1, \ldots, m_2 - 1$, we have

$$m_1m_2(i, j) = \sum_{i',\, j',\, u_1,\, u_2} (i', j') \alpha_1^{u_1(i' - i)} \alpha^{u_2(j' - j)}, \tag{7}$$

for when we carry out the summation on the right with respect to the $u$'s every term vanishes except when $i' = i \pmod{m_1}$ and $j' = j \pmod{m_2}$ simultaneously, since $\sum_{u_1} \alpha_1^{u_1k} = 0$, $k \not\equiv 0 \pmod{m_1}$ and it equals $m_1$ for $k = 0$, with a similar result involving $\alpha_2$. Using (4), consider the relation

$$\sum_{i,j} m_1^2m_2^2(i, j)(i + h, j + k) = \sum_{u_1,\, u_2,\, u_1',\, u_2'}^{i,\, j} \psi(\alpha_1^{u_1}, \alpha_2^{u_2}) \psi(\alpha_1^{u_1'} \alpha_2^{u_2'}) \cdot \alpha_1^{-iu_1 - iu_1' - hu_1'} \alpha_2^{-ju_2 - ju_2' - ku_2'}. \tag{7a}$$

Carrying out the summation on the right we find that every term vanishes except when $u_1 + u_1' \equiv 0 \pmod{m_1}$ and $u_2 + u_2' \equiv 0 \pmod{m_2}$ simultaneously, hence the right-hand member reduces to $m_1m_2$ times the expression

$$\sum_{u_1, u_2} \psi(\alpha_1^{u_1}, \alpha_2^{u_2})\psi(\alpha_1^{-u_1}, \alpha_2^{-u_2})\alpha_1^{hu_1}\alpha_2^{ku_2}. \tag{8}$$

Now as in the proof of Theorem I of $N$, we consider the five cases:

$$\begin{array}{rl} \text{I.} & \alpha_1^{u_1} = 1, \alpha_2^{u_2} = 1. \\ \text{II.} & \alpha_1^{u_1} = 1, \alpha_2^{u_2} \neq 1. \\ \text{III.} & \alpha_1^{u_1} \neq 1, \alpha_2^{u_2} = 1. \\ \text{IV.} & \alpha_1^{u_1}\alpha_2^{u_2} = 1, \alpha_1^{u_1} \neq 1, \alpha_2^{u_2} \neq 1. \\ \text{V.} & \alpha_1^{u_1}\alpha_2^{u_2} \neq 1, \alpha_1^{u_1} \neq 1, \alpha_2^{u_2} \neq 1. \end{array}$$

For Case I we find that (8) reduces to $(p^n - 2)^2$. For Case II, we find since $\psi(1, \alpha_2^{u_2}) = -1$ for $\alpha_2^{u_2} \neq 1$, the value $\sum\limits_{u_2} \alpha_2^{ku_2}$ and similarly for Case III. In Case IV, using the relation $\alpha_1^{u_1}\alpha_2^{u_2} = 1$ as in the proof of (15) of $N$, we find that the corresponding terms reduce to $\sum\limits_{v=1}^{d-1} \alpha_1^{(h-k)\,2s_1}$, where $d$ is the G. C. D. of $m_1$ and $m_2$ and $m_1 = s_1d$.

For Case V we find that the corresponding terms become

$$p^n\Big(\sum_{\substack{u_1 \neq 0 \\ u_2 \neq 0}} \alpha_1^{hu_1}\alpha_2^{ku_2} - \sum_{v=1}^{d-1} \alpha_1^{(h-k)s_1v}\Big).$$

Hence summing the terms for Cases I to V, inclusive, gives from

$$m_1m_2\sum_{i,j}(i, j)(i + h, j + k) = (p^n - 2)^2 + \sum_{u_2 \neq 0} \alpha_2^{ku_2} + \sum_{u_1 \neq 0} \alpha_1^{hu_1} + \sum_{v=1}^{d-1} \alpha_1^{(h-k)vs_1} + p^n\Big(\sum_{\substack{u_1 \neq 0 \\ u_2 \neq 0}} \alpha_1^{hu_1}\alpha_2^{ku_2} - \sum_{v=1}^{d-1} \alpha_1^{(h-k)vs}\Big). \tag{9}$$

This relation may be expressed, in terms of rational integers only, as follows:

Set

$$\begin{aligned} m_1m_2(i, j) &= V \text{ and } V(i, j) = S_1 \\ V(i, j + k) &= S_2 \text{ if } \alpha_1^{ks_1} \neq 1, \\ V(i, j + k) &= S_3 \text{ if } \alpha_1^{ks_1} = 1, \\ V(i + h, j) &= S_4 \text{ if } \alpha_1^{hs_1} \neq 1, \\ V(i + h, j) &= S_5 \text{ if } \alpha_1^{hs_1} = 1, \\ V(i + h, j + k) &= S_6 \text{ if } \alpha_1^{s_1(h-k)} \neq 1, \\ V(i + h, j + k) &= S_7 \text{ if } \alpha_1^{s_1(h-k)} = 1, \end{aligned}$$

where in each of the above $h \not\equiv 0 \pmod{m_1}$, $k \not\equiv 0 \pmod{m_2}$. We find by the use of (9),

$$
\begin{aligned}
S_1 &= (p^n - 2)^2 + m_1 + m_2 + d - 3 + p^n((m_1 - 1)(m_2 - 1) - d + 1) \\
&= (p^n - 1)^2 - (p^n - 1)(m_1 + m_2 + d) + p^n m_1 m_2, \qquad (10)
\end{aligned}
$$

and similarly

$$S_2 = (p^n - 1)^2 - (p^n - 1)m_1, \qquad (11)$$

$$S_3 = (p^n - 1)^2 - (p^n - 1)(m_1 + d), \qquad (12)$$

$$S_4 = (p^n - 1)^2 - (p^n - 1)m_2, \qquad (13)$$

$$S_5 = (p^n - 1)^2 - (p^n - 1)(m_2 + d), \qquad (14)$$

$$S_6 = (p^n - 1)^2, \qquad (15)$$

$$S_7 = (p^n - 1)^2 - d(p^n - 1). \qquad (16)$$

Hence we have

THEOREM I. *Let $(i, j)$ represent the number of sets $s$, $t$ satisfying (3) with $i$ and $j$ fixed. $s$ in the range $0, 1, \ldots, m_1' - 1$, $t$ in the range $0, 1, \ldots, m_2' - 1$, $g$ being a primitive root of the finite field $F(p^n)$, $p$ prime, then the relations (10) to (16), inclusive, hold with $m_1$ and $m_2$ divisors of $p^n - 1$, $d$ the G. C. D. of $m_1$ and $m_2$, with $m_1 = s_1 d$.*

In addition to expressing $S_1$ as the sum of squares in (10) we may obtain other expressions from the above theorem involving sums of squares. For example we have

$$\sum_{i,j}((i, j) - (i + h, j + k))^2 = 2S_1 - 2S_6 \qquad (17)$$

where $\alpha_1^{s_1(h - k)} \neq 1$.

We now consider some other types of quadratic relations between numbers of solutions of (1) with the particular case when $m_1 = m_2 = m$. Let $\alpha = e^{2i\pi/m}$ and set

$$a_{s,r} = \sum_{\rho = 0}^{m-1} (\rho, r + s\rho).$$

We shall show how to find an explicit expression for

$$\sum_{r=0}^{m-1} a_{s,r} a_{s,r+k} = A.$$

To effect this we employ the relation (6) and (7a) for $m_1 = m_2 = m$. We have

$$A = \sum_{\substack{\rho', \mu_1, \mu_2 \\ \mu_1', \mu_2' \\ r}} \psi(\alpha^{\mu_1}, \alpha^{\mu_2})\psi(\alpha^{\mu_1'}, \alpha^{\mu_2'})\cdot\alpha^{-\rho\mu_1 - (r + s\rho)\mu_2 - \rho'\mu_1' - (r + s\rho' + k)\mu_2'} \qquad (18)$$

where all the quantities involved in the summation on the right range, independently over 0, 1, ..., $m - 1$. The second factor in the right-hand member of this can be written in the form

$$A = \sum_{\substack{\mu_1\mu_2 \\ \mu_1'\mu_2'}} \psi(\alpha^{\mu_1}, \alpha^{\mu_2})\psi(\alpha^{\mu_1'}, \alpha^{\mu_2'})\cdot\sum_{\rho}\alpha^{-\rho(\mu_1 + s\mu_2)} \sum_{\rho'}{}^{\rho'(\mu_1' + s\mu_2')}\sum_{r}\alpha^{r(\mu_2 + \mu_2')} \tag{19}$$

which reduces to zero unless

$$\begin{aligned} \mu_1 + s\mu_2 &\equiv 0 \\ \mu_1' + s\mu_2' &\equiv 0 \\ \mu_2 + \mu_2' &\equiv 0 \end{aligned}$$

modulo $m$, simultaneously. Hence

$$A = \sum_{\mu_2} m^3\psi(\alpha^{-s\mu_2}\alpha^{\mu_1})\psi(\alpha^{s\mu_2}, \alpha^{-\mu_1})\alpha^{k\mu_2}$$

which in view of (6), (7), and (7a), and the known values of each $\psi$ when $\alpha^{\mu_2} = 1$, we obtain (A. Hurwitz[4])

$$\sum_{r=0}^{m-1} a_{sr}a_{s,\, r+k} = (p^n - 4)m'; \tag{20}$$

$$\sum_{r=0}^{m-1} a_{sr}^2 = (p^n - 4)m' + p^n, \tag{21}$$

with $s$ and $s - 1$ each prime to $m$, $k \not\equiv 0 \pmod{m}$, $mm' = p^n - 1$.

From (20) and (21) we obtain

$$\sum_{r=0}^{m-1} (a_{s,\, r} - a_{s,\, r+k})^2 = 2p^n \tag{22}$$

which gives $m^2$ representations (possibly not all different) of $2p^n$ as the sum of $m$ squares. Similarly (17) gives analogous representations of various types of integers.

It is easy to see that we may generalize (20) to include the cases when $s$ or $s - 1$ may not be prime to $m$.

---

[1] These PROCEEDINGS, **33**, 236–242 (1947); **34**, 62–66 (1948).

[2] *Ibid.*, **35**, 481–487 (1949). However, that paper (which will be referred to as $N$) employs a classification of various equations of the type (1) which we shall use here.

[3] These PROCEEDINGS, **32**, 317–319 (1946). Davenport and Hasse, *Jour. für die Mathematik* (*Crelle*), **172**, 174 (1935) used an analogous expression involving characters and the number of solutions of (2) without the restriction $n_1n_2 \neq 0$.

[4] *Jour. für die Mathematik* (*Crelle*), **136**, 390 (1909) for $n = 1$ and $m$ an odd prime. The success of Hurwitz' method appears to depend on the fact that he restricts $m$ in this way. His argument is quite long.

# ON THE USE OF INDETERMINATES IN THE THEORY OF EXPONENTIAL SUMS

By H. S. Vandiver

Department of Applied Mathematics, University of Texas

Communicated October 31, 1949

Consider the sum

$$\psi(\alpha) = \sum_{a} \alpha^{\mu_1 \operatorname{ind} a + \mu_2 \operatorname{ind} (1 - a)}, \tag{1}$$

where $a$ ranges over all the elements of a finite field $F(p^n)$, of order $p^n$ with $p$ an odd prime, $\mu_1$, $\mu_2$ and $\mu_1 + \mu_2$ are integers, none divisible by $m$, $\alpha = e^{2i\pi/m}$, ind $a$ is given by $g^{\operatorname{ind} a} = a$, where $g$ is a primitive root of $F(p^n)$, $\alpha^{\operatorname{ind}(0)} = 0$, $p^n - 1 \equiv 0 \pmod{m}$. Many relations[1] have been given involving equation (1) but few of these express all that may be inferred from the proofs given for them. But to have given the whole story would have led to quite complicated formulas in many cases, so to illustrate the points we have in mind we start our discussion with some of the simplest relations along these lines.

It is known that

$$\psi(\alpha)\psi(\alpha^{-1}) = p^n, \tag{2}$$

if each of $\mu_1$, $\mu_2$, $\mu_1 + \mu_2 \not\equiv \pmod{m}$. If we seek the equivalent of this using indeterminates, we note that we can write

$$\psi(x)\psi(x^{m-1}) = p^n + V(x)F(x), \tag{3}$$

where $F(x)$ is the irreducible polynomial of degree $\phi(m)$, the indicator of $m$, with rational integral coefficients which vanishes for all the primitive roots of $x^m = 1$, and $V(x)$ is a polynomial with rational integral coefficients. The relation (3) holds since

$$\psi(x)\psi(x^{m-1}) - p^n \tag{3a}$$

vanishes for any root of $F(x) = 0$ and therefore is a polynomial which is a multiple of $F(x)$. Now for $m$ composite $F(x)$ can be quite a complicated polynomial. *But it is possible*[2] *to prove* (2) *so that the only property of* $\alpha$ *employed is* $\alpha^{m-1} + \alpha^{m-2} + \ldots + 1 = 0$. Hence we are entitled to state in place of equation (3),

$$\psi(x)\psi(x^{m-1}) = p^n + W(x)\frac{x^m - 1}{x - 1}, \tag{4}$$

since equation (3a) is a polynomial which vanishes for all the roots, $\neq 1$, of $x^{m-1} + x^{m-2} + \ldots + 1 = 0$. Now we first give a very simple example to show that (4) is more powerful than (3). If we substitute $x = 1$ in equation (3), we obtain, if we let $m = q^r$, $r > 1$, $q$ prime

$$\psi(1)^2 = p^n + qW(1) \equiv p^n \pmod{q},$$

whereas if we substitute $x = 1$ in equation (4), we have $x(1)^2 \equiv p^n \pmod{q^r}$.

Among the things we can do with such identities as equations (3) and (4) is to differentiate them. Starting with equation (3), we have for $m = q^2$, differentiating with respect to $x$, calling $A(x)$ the ordinary derivative of each member and setting $x = 1$ in the result,

$$A(1) \equiv qV'(1) \pmod{q^2}.$$

The same procedure with equation (4) gives

$$A(1) \equiv 0 \pmod{q^2}.$$

Hence equation (4) has again given us more information than (3).

The above remarks show that when $m$ is prime the relation (3) is the same as (4) so that the use of an indeterminate yields nothing more than we obtain from equation (3).

The form (1) is an expression associated with cyclotomy but the writer has noted that many other types of exponential sums, apparently not related to this topic, have been considered by various writers and results derived concerning them *with the use of the single property*

$$1 + \alpha + \ldots + \alpha^{m-1} = 0,$$

*if* $m$th *roots of unity are employed.* Hence our observations will apply to these latter types also. On the other hand, such questions as the factorization into ideals of $(\psi(\alpha))$ makes use of the fact that the degree of the field is $\phi(m)$, the indicator of $m$, so that we may expect nothing new, by the use of our methods, along this line.

Before proceeding to the proof of some more general results we shall explain the idea of formal exponential differentiation[3] which we wish to employ. We first consider a polynomial $f(x)$ in $x$ with coefficients in $R$, where $R$ is the ring of integers in $K(\gamma, \alpha)$ an algebraic field including $\alpha$, or considering $x$ as an integral transcendental adjunction, we consider a quantity in $R[x]$. We first find the ordinary derivative $f'(x)$ of $f(x)$ and then multiply by $x$ defining the formal exponential derivative, with respect to $x$, as $xf'(x)$. We may prove easily the relation, if $A$ and $B$ are in $R[x]$,

$$\bar{D}_n(AB) = \sum_{i=0}^{n} \binom{n}{i} \bar{D}_i(A)\bar{D}_{n-i}(B). \qquad (5)$$

We employ as a definition

$$\bar{D}\left(\frac{F}{G}\right) = \frac{G\bar{D}(F) - F\bar{D}(G)}{G^2}, \tag{6}$$

where $F$ and $G$ are in $R[x]$.

We now prove

THEOREM I. *Let* $\alpha = e^{2i\pi/m}$ *and*

$$f(x) = \theta_r x^r + \theta_{r-1}x^{r-1} + \ldots + \theta_0,$$

*where the* $\theta$'*s are in the ring* $R$ *of integers in* $K[\gamma, \alpha]$, *an algebraic field including* $\alpha$. *Suppose that* $f(x) = 0$ *for* $x = \alpha, \alpha^2, \ldots, \alpha^{m-1}$ *and* $g(x)$ *is another polynomial in* $R[x]$, *assuming the ideal* $(g(\alpha^u), m) = (1)$ *in* $R$ *with* $u$ *an integer such that if* $\alpha^{uk} = 1$, *then* $k$ *contains at least two distinct prime factors. Then*

$$\left[\bar{D}_s\left(\frac{f(x)}{g(x)}\right)\right]_{x = \alpha^u} \equiv 0 \pmod{m}. \tag{7}$$

In the above, $\bar{D}_s$ means that the fraction $f(x)/g(x)$ is differentiated $s$ times using formal exponential differentiation. For the proof we have since $f(x)$ vanishes for $x = \alpha, \alpha^2, \ldots, \alpha^{m-1}$,

$$f(x) = \frac{x^m - 1}{x - 1}V(x),$$

with $V(x)$ in $R[x]$ and

$$\frac{f(x)}{g(x)} = \frac{x^m - 1}{x - 1}\cdot\frac{V(x)}{g(x)}. \tag{8}$$

Using equation (5) we have

$$\bar{D}_t\left(\frac{x^m - 1}{x - 1}\right) = \sum_{i=0}^{t} \bar{D}_i\left(\frac{1}{x - 1}\right)\bar{D}_{t-i}(x^m - 1). \tag{9}$$

We note also that for $i < t$

$$\bar{D}_{t-i}(x^m - 1) = m^{t-i}x^m,$$

which for $x = \alpha^u$ reduces to an integer divisible by $m$, as well as

$$[x^m - 1]_{x = \alpha^u} = 0.$$

Also if we express, after employing equation (6),

$$\bar{D}_i\left(\frac{1}{x - 1}\right)$$

as a simple fraction in its lowest terms, the denominator is a power of $x - 1$. When we substitute $x = \alpha^u$ in $(x - 1)^k$ we obtain a unit[4] in

$K(\gamma, \alpha)$ by the hypothesis concerning $\alpha^u$. Now examine

$$\bar{D}_v\left(\frac{V(x)}{g(x)}\right)$$

and differentiate it as the product of $V(x)$ and $(g(x))^{-1}$ as in equation (9). The result may be expressed as a simple fraction whose denominator is a power of $g(x)$, and by the hypothesis concerning $g(x)$ given in our theorem the left-hand member of equation (7) may be reduced to a fraction of the form

$$\frac{\omega}{\beta},$$

with $\omega$ and $\beta$ in $R$ and $\omega \equiv 0 \pmod{m}$, $(\beta, m) = (1)$. This proves relation (7).

In applying Theorem I it is clear that $f(x)$ does not have to be given explicitly as a polynomial; we may, for example, work with $f(x)$ expressed as a product of polynomials.[5]

We do not have

$$\left[\bar{D}_t\left(\frac{x^m - 1}{x - 1}\right)\right]_{x=1} \equiv 0 \pmod{m}$$

for all values of $t$, but we may extend Theorem I to include $x = 1$, under a certain assumption concerning $f(1)$. For in equation (1) let, if $V_1$ is in $R[x]$,

$$V(x) = (x - 1)V_1 + \delta,$$

where $\delta$ is in $R$. Substitution in equation (1) gives

$$f(x) = (x^m - 1)V_1(x) + \delta\frac{x^m - 1}{x - 1}.$$

Assume $f(1) \equiv 0 \pmod{m^2}$ we obtain by setting $x = 1$,

$$f(1) \equiv 0 \equiv \delta m \pmod{m^2},$$

or $\delta = m\tau$, with $\tau$ in $R$. It is then easy to see that Theorem I is also true if $\alpha^u$ is replaced by unity if we add the assumption

$$f(1) \equiv 0 \pmod{m^2}.$$

For other cases which do not come under the assumption of Theorem I, we note that if we have $m = l$ a prime and $f(x) = 0$ for $\alpha, \alpha^2, \ldots, \alpha^{l-1}$, then

$$\frac{f(x)}{g(x)} = \frac{x^l - 1}{x - 1}\cdot\frac{V(x)}{g(x)},$$

also

$$D_i\left(\frac{x^l - 1}{x - 1}\right) = x + 2^i x^2 + \ldots + (l - 1)^i x^{l-1}.$$

We have the identical congruence

$$1 + x + x^2 + \ldots + x^{l-1} \equiv (x - 1)^{l-1} \pmod{l},$$

whence, employing equation (5)

$$x + 2^i x^2 + \ldots + (l - 1)^i x^{l-1} \equiv (x - 1)^{l-i-1} h(x) \pmod{l},$$

for some $h(x)$ in $R[x]$ and this shows that for $i < l - 1$ and the restriction on $g(x)$ given in Theorem I we have

$$\left[D\left(\frac{f(x)}{g(x)}\right)\right]_{x=\alpha} \equiv 0 \pmod{\lambda^{l-i-1}}, \tag{10}$$

where $\lambda = (1 - \alpha)$.

---

[1] Cf. for example for $n = 1$, Bachmann, P., *Die Lehre von der Kreisteilung*, ed. 2, Leipzig and Berlin, Teubner (1921), pp. 88–93, 122–150; Weber, H., *Lehrbuch der Algebra*, Braunschweig, Vieweg, (1895), ed. 1, vol. 1, 564–574, vol. 2, 632–638; for $n$ general, Mitchell, H. H., *Trans. Am. Math. Soc.*, **17**, 167 (1916), *Ann. Math.* II, **18**, 120 (1917); Davenport, H., and Hasse, H., *J. f. Math.* (Crelle), **172**, 152–154 (1935) and other references in the Report of the Committee on Algebraic Numbers, National Research Council, *Bull. Nat. Research Council*, U. S., pt. 3, No. 28, 50–53, Washington (1923).

[2] Mitchell, H. H., *loc. cit.*, first reference.

[3] These PROCEEDINGS, **28**, 24–27 (1942).

[4] Hilbert, D., *Gesammelte Abhandlungen* Bd. 1, Berlin, Springer (1932), pp. 199, 203.

[5] If $l$ is an odd prime dividing $m$ then the method of introducing an indeterminate and obtaining congruences, modulo $l$, from relations involving exponential sums has been employed in various connections. For example, Kummer proceeded in this way in obtaining the so-called Kummer criteria for the solution of

$$x^l + y^l + z^l = 0$$

in integers (cf. Vandiver, *Ann. Math.* II, **21**, 73–78 (1919) for a proof in modern form). Further (these PROCEEDINGS, **34**, 197–203 (1948)) such devices were employed by the writer to obtain congruences involving the $l$th power character

$$\left(\frac{1 - \alpha}{\mathfrak{p}}\right)_l \equiv (1 - \alpha)^{(N(\mathfrak{p}) - 1)/l} \pmod{\mathfrak{p}},$$

where $\alpha = e^{2i\pi/m}$, $m \equiv 0 \pmod{l}$, $\mathfrak{p}$ is a prime ideal in $K(\alpha)$, and $N(\mathfrak{p})$ is the norm of $\mathfrak{p}$. Also the writer, *Trans. Am. Math. Soc.*, **32**, 401 (1929), lemma II, used similar methods when $m = l$ to find an explicit expression for

$$\left(\frac{1 - \alpha}{\mathfrak{p}}\right)_l,$$

when $\mathfrak{p}$ is *any* prime ideal in $K(\alpha)$.

# *SOME RECENT APPROACHES TO THE EXPERIMENTAL STUDY OF HUMAN FATIGUE*

BY LEONARD CARMICHAEL, JOHN L. KENNEDY AND LEONARD C. MEAD

TUFTS COLLEGE, MEDFORD, MASSACHUSETTS

Read before the Academy, October 25, 1949

More than a quarter of a century ago it was proposed that the word fatigue be banished from the vocabulary of science.[8] As the title of this paper attests, this sentence of expulsion has never been enforced. The reason that the concept of fatigue retains its place in scientific writing is not hard to seek. Everyone personally feels that he knows what it is to be tired or fatigued. But in spite of this common-sense understanding, a rigorous definition acceptable to all workers in this field has never been devised. Sometimes today the word fatigue is used to apply to subjective feelings. More often it is used to describe objectively measurable changes in behavior or in underlying physiological mechanisms which follow continued activity.

Most scientists who have not themselves studied fatigue probably think of it in relation to characteristic work curves. These textbook curves ordinarily show with continued work a gradual decrement, as developed in the classical work of Mosso[7] with the ergograph. Unfortunately a transfer of the idea of this continuous work decrement to an understanding of the behavior of the total living organism in normal life situations is often most misleading.

The first of the experimental approaches to the study of fatigue in the normal adult human individual to which we wish to make brief reference in this paper is an experiment conducted by Carmichael, Dearborn, Hoffman and their associates on the continuous recording of the eye movements during long periods of continuous reading.[1]

The most important conclusion drawn from these experiments is that if subjects are sufficiently highly motivated they can continue to read with undiminished comprehension for six hours. During this whole period also there was no statistically significant change in any detail of the many movements of the eyes which are required in reading. It should be noted that while these reading experiments seemed to show clearly that 40 subjects could read for six hours without a work decrement, a less highly motivated group of subjects in a previous experiment had begun to show lapses in continuous reading in a four-hour period of work.

Other experiments may now be mentioned which were conducted several years ago under a National Defense Research Committee contract under the direction of Mead.[2, 3, 6, 11–14] The general purpose of these studies was to determine the conditions under which a decrement in behavior could be measured in subjects performing certain military jobs. The typical performances studied were stereoscopic ranging and eye-hand tracking. Measurements of other types of performance were also included. The usual precautions which are maintained in modern fatigue research were observed, including the conduct of preliminary practice sessions prior to the experiment proper so given as to eliminate improvement through learning.

Four subjects were required to follow a 74-hour schedule broken as follows: 42 hours on duty, which consisted of visual lookout, stereoranging or tracking; 16 hours sleeping; 8 hours eating; and 8 hours at liberty.[11] The periods during which these activities were engaged in were irregular. The stressful feature of this experiment was the irregularity and unpredictability of routines. The results showed no consistent decrement in visual lookout performance (number of target lights reported), in stereoprecision or stereoconsistency scores, or in tracking.

Despite these objective indications of lack of fatigue, the observers complained of increasing feelings of fatigue as the test progressed, particularly during the last duty period from 4:00 A. M. to 12:00 noon after $2^1/_2$ hours' sleep. Some social discord between subjects was noted.

Another experiment intended to induce a greater amount of physiological fatigue was arranged.[12] In this experiment subjects began work at 2:00 P. M. Friday and action was continued without sleep through 11:00 A. M. the following Sunday. Sets of visual ranging and tracking tests were again given intermittently during this period as in the previous test. In order to sustain interest in good stereoperformance during the trials, bonuses of 40 cents were paid on the first four trials if any precision score was better than that obtained during preliminary practice. During the subsequent nine sets of experimental trials, a bonus was paid if a score was better than that obtained during the first four experimental trials. The objective measures of fatigue were comparisons of performances during the first and the last 24 hours. Other observations and records that were taken during the course of the experiment were eye movements during ranging and tracking, accuracy and time to complete multiplication problems, reading rate and comprehension, and hand steadiness. An attempt was made to measure changes in group morale by observing behavior during card games scheduled at three stages of the three-day period. As in the previous experiments reported here, objective evidences of fatigue were negative in these tests. Stereoprecision and consistency scores during the first and last 24 hours showed no statistically significant differences. Tracking accuracy actually increased as the test progressed. This finding may perhaps be attributed to insufficient pre-test practice. Eye movement records and the speed and accuracy of multiplication showed no significant change. Hand steadiness increased during the first

28 hours and then remained constant. Reading speed showed a drop on the last two tests on Sunday morning, but reading comprehension maintained a constant level throughout. Activity, talkativeness and sociability declined during the first 24 hours. The second card game on Saturday afternoon was done quietly and without much interest. Social communication dwindled still more after this second game. A game scheduled for Sunday morning was cancelled because the subjects seemed so unresponsive. No important indications of irritability or discontent were observed. Toward the end of the experiment conversation, when it occurred, was centered on how, when and how long the subjects were going to sleep when the experiment was over. These appearances of sluggishness tended to disappear during meals and during the taking of the various tests. Again it was notable that no performance decrement was apparent in behavior which was objectively measured even after such long periods of sleep deprivation.

In another study an additional attempt was made to induce objective and measurable signs of fatigue by increasing the duress of the experimental conditions.[13] Ten male subjects were now not only required to undergo a sleep deficit but also to carry out protracted physical exercise. The test battery consisted of the following measures: (1) mean and standard deviation scores of stereorange settings, (2) differential brightness matching, (3) complex visual reaction time, (4) the average time per minute when the tracking error exceeded an arbitrary value, (5) the number of words correctly crossed out in a persistence test. This latter test consisted of crossing out all words containing both an "a" and a "t" on pages of Spanish, printed solid. There were 30 pages of Spanish in all and the subject worked for three minutes on each page for a total of 90 minutes. Bonus payments were again made for exceptional performance. A week's preliminary practice was given on all tests except the persistence test. The pre-test baseline for the persistence test occurred on the evening before the experiment proper began. The visual tests were now administered before and after three 10-mile walks so that the subjects took the tests four times and walked a total of 30 miles.

With respect to stereoscopic ranging, eye-hand tracking, differential brightness discrimination, and complex visual reaction time, no statistically significant difference between behavior when starting the tests and after sleeplessness and exercise was found. A statistically significant decrement was found in the persistence tests before and after the long hike. The performance on this test after the exercise was superior at the outset of the 90-minute work period to that found before the exercise. In other words, the subjects were physically capable of performing this monotonous task after exercise but were either unable or unwilling to exert themselves to maintain this level of performance. Toward the end of the experiment and after the 30-mile walk the subjects almost literally "drooped" in the laboratory. Yet on the four different visual tests they pulled themselves together for the time required to make the necessary responses so that their performance was still in general as satisfactory as at first.

These experiments so far seem to suggest that a performance decrement is unlikely to be demonstrated on tasks which require discrete and short-time periods for their execution.[14] In all these experiments, including a comparison of the two reading experiments noted above, it seems that decrement in performance should be looked for with respect to an individual's interest, willingness, effort or motivation to perform a given task. If the task is continuous and of long duration, it is likely that the first existence of a physiological work decrement will be found in a decline in the subject's desire to make himself do the task especially for long-continued periods of time.

Another experiment was conducted to test the generalizations just made.[2,3] In this study a central eye-hand pursuit task had to be performed continuously while attention and occasional response had to be given simultaneously to a number of peripheral subtasks. The time for correctly performing any one of these tasks was recorded on separate clocks. Whenever either the central task or one of the subtasks was not performed adequately, the moving target would stop its motion. The total time to complete this complex task was the main criterion of performance.

Nine male college students took this test periodically throughout a three-day period in which they were deprived of sleep. Measurements of stereo-ranging were also taken at intervals during the three days. Monetary bonuses were again used to keep motivation high. The total-time scores required to complete the complex task showed a significant decrement between the first and third days. But, as expected, no deficit in stereo-ranging was found.

These results support the conclusion of the previous experiments, namely, that performance decrements due to physiological stress may be expected to occur in tasks which require continuous and prolonged effort. The fact that a decrement did occur in performing the complex task after the loss of 50 hours of sleep raises the interesting question as to whether the decline was due to actual physiological incapacity or to a breakdown of the willingness of the subjects to expend effort to keep the performance up to standard.

Because of the importance of "motivation" in these experiments and for other reasons, Kennedy, Travis and their associates developed a sensitive measure of muscle tonus which could be recorded continuously.[4, 5, 9, 10] They were interested in discovering whether or not this measure might be a valid indicator of drowsiness or loss of motivation or lapses in alertness; that is, they were attempting to find a physiological process of which records could be made and which would parallel the fluctuations in a continuous performance, previously noted in this paper as characterizing the onset of fatigue as measured in the behavior of the total organism during long-continued periods of activity.

Muscle fibers when they contract produce characteristic potential changes. After much exploration it was found that if electrodes were placed on the skin over the frontalis muscle just above the eyebrows records could be secured that were significantly related to the alertness of the subjects studied.

A 4-channel electroencephalograph, with associated pre-amplifiers and power amplifiers, electronic counters, a voltage discriminator for counting spikes above a given amplitude, and an integrator or accumulator for continuously integrating the output from the muscle each second, was assembled.

This device has now been made in a portable form and tested in submarines, trucks and aircraft. In general, the results secured with this alertness indicator indicate that there is a definite quantitative relationship between the electric output of the chosen muscle group and the level of alertness of the subject. The highest alertness is associated with maximal muscle potential output and this output fluctuates in relation to the state of the subject until the lowest output in sleep is secured.

In conclusion then it may be said of the experiments briefly summarized here that they agree in the showing that so-called "fatigue" as it presents itself in the performance of the total normal adult human being has certain special characteristics. When the task required is not beyond the subject's normal ability, but when it is continued for a long time, the change in performance first noted is in apparently involuntary lapses in an otherwise continuous and effective performance. As such continuous and monotonous tasks go on there seems to be a conflict between motivation, or the desire to continue, and level of performance. When the subject sets himself to carry out a skilled performance after even a long period of interpolated physical work he is typically able to perform as well as in a rested state. In general, however, as time without rest progresses the number and duration of lapses increase until in a typical case sleep ensues. The pattern of human fatigue as given in these experiments is thus seen to be one of periodic blocks or interruptions in performance of varying frequency and duration rather than the continuous decrement as shown in the work curves secured in the study of isolated muscle groups.

---

[1] Carmichael, L., and Dearborn, W. F., *Reading and Visual Fatigue*, Boston, Houghton Mifflin Co., 1947.

[2] Clark, R. E., and others, "The Effects of Sleep Loss on Performance of a Complex Task" (OSRD, 1943; Publ. Bd., No. 20286), Washington, D. C., U. S. Dept. Commerce, 1946, p. 19.

[3] Hoffman, A. C., and Mead, L. C., "The Performance of Trained Subjects on a Complex Task of Four Hours Duration" (OSRD, 1943; Publ. Bd., No. 20284), Washington, D. C., U. S. Dept. Commerce, 1946, p. 16.

[4] Kennedy, J. L., and Travis, R. C., "Prediction and Control of Alertness. II. Continuous Tracking," *J. Comp. & Physiol. Psychol.*, **41**, 203–210 (1948).

[5] Kennedy, J. L., and Travis, R. C., "Prediction of Speed of Performance by Muscle Action Potentials," *Science*, **105**, 410–411 (1947).

[6] Mead, L. C., "Research and Development Work: Summary report from Aug. 1, 1942 to July 1, 1943" (OSRD, 1943; Publ. Bd., No. 20830), Washington, D. C., U. S. Dept. Commerce, 1946, p. 15.

[7] Mosso, A., *Fatigue* (translated by M. Drummond and W. B. Drummond), New York, Putnam, 1904.

[8] Muscio, B., "Is a Fatigue Test Possible? (A report to the Industrial Fatigue Research Board)," *Brit. J. Psychol.*, **12**, 31–46 (1921–1922).

[9] Travis, R. C., and Kennedy, J. L., "Prediction and Automatic Control of Alertness. I. Control of Lookout Alertness," *J. Comp. & Physiol. Psychol.*, **40**, 457–461 (1947).

[10] Travis, R. C., and Kennedy, J. L., "Prediction and Control of Alertness. III. Calibration of the Alertness Indicator and Further Results," *Ibid.*, **42**, 45–57 (1949).

[11] Tufts College, "Fatigue Tests: Three-Day Test of Fatigue Under Conditions of Long Hours on Duty, Limited Sleep" (OEMsr-581, Rep. 3, 1942; Publ. Bd. No. L77776), Washington, D. C., U. S. Dept. Commerce, 1947, p. 22.

[12] Tufts College report to National Defense Research Comm. under Contract OEMsr-581, Report No. 4, "Effect of Sleep Deprivation Upon Performance," 1942, p. 17.

[13] Tufts College report to National Defense Research Comm. under Contract OEMsr-581, Report No. 6, "Effect of a Thirty Mile Hike on Stereoranging, Tracking, and Other Tasks, 1943, p. 30.

[14] U. S. National Defense Research Committee, "The Effects of Motivation and of Fatigue on Stereoscopic Ranging and Direct Tracking," Report to the Services No. 56, Div. 7, Fire Control (OSRD, 1942, Publ. Bd., No. 34118), Washington, D. C., U. S. Dept. Commerce, 1946, p. 35.

---

# *ON THE INTERPRETATION OF MULTI-HIT SURVIVAL CURVES**

BY K. C. ATWOOD AND A. NORMAN†

DEPARTMENT OF ZOOLOGY, COLUMBIA UNIVERSITY

Communicated by E. N. Harvey, September 7, 1949

*Introduction.*—Survival curves, which comprise a large proportion of the data on the effects of radiation on microorganisms, generally fall into two categories: the exponential or single-hit and the sigmoid or multi-hit. It is unfortunate that in the interpretation of such data few criteria have been available other than the form of the curves themselves, but this alone has provided sufficient grounds for speculation on some properties of the organisms. Exponential survival can be explained either on the basis of a single hit to kill[1, 2] or a population distributed exponentially with respect to resistance to radiation.[3] Several considerations which make the single-hit hypothesis seem the more plausible have been pointed out.[2, 4] Lea[5] further proposed that lethal mutation might account for the bactericidal effects of radiation, a notion which arises naturally from the finding of single-hit gene mutations in higher organisms, but which has been neither supported nor disproved by experiment. Curves of the sigmoid type may be interpreted in a similar variety of ways. If the distribution hypothesis is rejected, the number of hits, $n$, required to produce some criterion of inviability can be estimated from the shape of the curve. On this basis, the values of $n$ for different types of radiation on the same material can be compared,[6] or some biological situation can be examined which may be expected to alter $n$.[7]

*Theoretical.*—Any determination of $n$ rests ultimately upon some assumption about the way in which the sublethal hits are accumulated. The hypotheses proposed to describe multi-hit curves reduce to two which differ in this basic assumption. The first and most often used of these may be derived as follows: Let $D$ be the average number of hits occurring within a unit volume, and let $a$ be the sensitive volume of the organism. Then $aD$ is the average number of hits within the sensitive volume $a$. If the hits occur independently and at random, the probability that $X$ hits fall within the volume $a$ is given by:

$$P_{(x)} = e^{-aD}\frac{(aD)^X}{X!}. \qquad (1)$$

If $n$ hits within $a$ are required to kill an organism, then all those having less than $n$ hits survive and the probability of survival is:

$$S = e^{-aD}\sum_{x=0}^{n-1}\frac{(aD)^X}{X!}. \qquad (2)$$

The assumption implicit in equation (2) is that the presence of one or more hits within the sensitive volume has no influence on the probability of obtaining further effective hits.

Since the second multi-hit hypothesis follows directly from the single-hit hypothesis, the latter will be briefly reiterated. Let $N$ be the number of organisms surviving a dose, $D$, of radiation. Let $N_0$ be the corresponding number at $D = 0$ and let $k$ be a constant independent of $D$. Then a single-hit curve is defined by:

$$-\frac{dN}{dD} = kN \qquad (3)$$

which integrated is

$$N = N_0e^{-kD} \qquad (4)$$

and the surviving fraction is

$$S = \frac{N}{N_0} = e^{-kD}. \tag{5}$$

From equation (5) it is seen that a plot of log fraction surviving against dose is a straight line of slope $-k$.[20]

Now consider a population of organisms each of which contains $n$ units such that the units are inactivated according to equation (5) and the presence of one or more viable units insures the viability of the organism. The probability that all the units in such a group of $n$ become inactivated is:

$$P_{(n)} = (1 - e^{-kD})^n. \tag{6}$$

Then the proportion of organisms surviving is:

$$S = 1 - (1 - e^{-kD})^n. \tag{7}$$

The assumption here is not simply that $n$ hits per organism are required, but that each of $n$ particles within the organism must be hit at least once. It is evident that equation (7) may also be derived as a special case of equation (2).[8] Expressions (2) and (7) describe curves of the same general form, and for low dose and high values of $n$ where terms of the order of $(kD)^n/n!$ can be ignored, they become identical. It is impossible in most cases to decide from experimental data which equation should be used as a working model. Nevertheless, the choice is important because it is possible for numerical values of $n$ obtained from one experiment to differ by an order of magnitude, depending upon which hypothesis is chosen. As will be seen later in this paper, there are additional criteria which point to expression (7), or some development thereof to cover additional assumptions, as the more reasonable in the interpretation of multi-hit curves.

The use of equation (2) and its extension to more complex cases has been discussed at length elsewhere,[9, 10] hence we confine ourselves to exploring some features of equation (7).

Expanding equation (7):

$$S = 1 - (1 - ne^{-kD} + \ldots \pm e^{-nkD}). \tag{8}$$

It is seen that as $D$ increases, the terms containing $e^{-2kD}$, $e^{-3kD}$, ..., etc., become negligible in comparison to $ne^{-kD}$. Thus, at high dose:

$$\log S = \log n - kD. \tag{9}$$

Extrapolating equation (9) to $D = 0$ gives $n$ (Fig. 1). Having obtained values for $n$ and $k$ we should be able to reconstruct the convex portion of the curve from equation (7) and show that the reconstruction fits the experi-

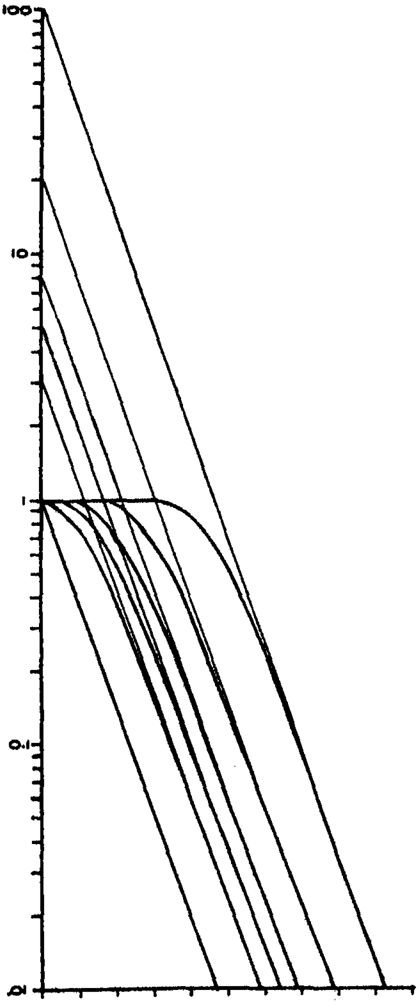


FIGURE 1

Ordinate, survival. Abscissa, dose in arbitrary units. Theoretical curves for $S = 1 - (1 - e^{-kD})^n$, $k = 1$. From left to right $n$ is 1, 3, 5, 8, 20 and 100. Linear portions of the curves are extrapolated to $D = 0$.

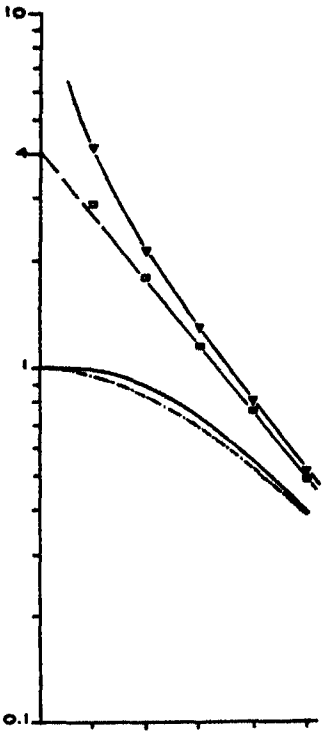


FIGURE 2

Ordinate, $S$ for lower curves, $g$ for upper curves. Abscissa, dose in arbitrary units. Lower curves: Solid line is theoretical curve for $S = 1 - (1 - e^{-kD})^n$ where $n = 4$ and $k = 0.43$. Broken line same, but $n$ has the distribution: 7% 1, 15% 2, 20% 3, 21% 4, 16% 5, 11% 6, 7% 7 and 3% 8, which averages 4. Upper curves: Triangles are points computed from solid line curve by formula: $g = -\ln(1 - S)$. Squares are points similarly computed from broken line curve. The $g$ plot for distributed $n$ is linear at higher survival than for constant $n$.

mental data in this region if equation (7) holds. However, while the basic assumption from which equation (7) is derived may hold, equation (7) may not be sufficient to describe experimental data because of at least two additional eventualities which seem, *a priori*, likely to be encountered. First, $n$ may not be a fixed number when $D = 0$, but instead $n$ might have some distribution in the population. Second, more than one process might be simultaneously occurring within each organism, each process having its own value of $n$ and $k$. In other words, there might be several classes of sensitive units with the presence of at least one viable member of each class necessary for survival.

In the case of a population distributed for $n$, equation (7) becomes:

$$S = 1 - [C_1(1 - e^{-kD})^{n_1} + C_2(1 - e^{-kD})^{n_2} + \ldots + C_i(1 - e^{-kD})^{n_i}], \qquad (10)$$

where $C_1 + C_2 + \ldots + C_i = 1$ and $C_i$ represents the fraction of the population having $n_i$ units. For high dose, as in equation (8), $(1 - e^{-kD})^n \cong 1 - ne^{-kD}$. With this substitution, equation (10) becomes:

$$S = e^{-kD}(C_1n_1 + C_2n_2 + \ldots + C_in_i). \qquad (11)$$

Thus:

$$\log S = -kD + \log \sum_{i=1} C_in_i = \log \bar{n} - kD \qquad (12)$$

where $\bar{n} \equiv \sum_{i=1}^{i} C_in_i$ is the ordinary arithmetical average. Thus extrapolation to $D = 0$ of the linear portion of equation (10) on a semi-log plot gives the average value of $n$ for the distribution present at $D = 0$.

The shape of the curve will be changed according to the type of distribution, since the presence of fractions of the population having values of $n$ below the average value will cause the curve to fall off more rapidly than it would when $n$ is a constant (Fig. 2, lower curves). Such changes are difficult to interpret by inspection or curve fitting, but can be perceived more easily from a semi-log plot of $-\ln(1 - S)$ against dose (Fig. 2, upper curves). The use of this device is justified as follows:

From equation (5), let $p = e^{-kD}$ be the probability that a unit (within an organism) survives. Let $P$ be the probability that $m$ units survive within an organism containing a total of $n$ units. Then:

$$P_{(m=i)} = \binom{n}{i} p^i (1 - p)^{n-i} \qquad (13)$$

As $D$ becomes large, $p$ approaches 0 and,

$$P_{(m=i)} \cong e^{-g}\frac{g^i}{i!} \tag{14}$$

where

$$g = np = ne^{-kD}. \tag{15}$$

The probability that an organism is killed is the probability that no units survive, and where equation (14) holds is:

$$P_{(m=0)} = e^{-g} = e^{-ne^{-kD}}. \tag{16}$$

Thus $e^{-g}$ is equated to the fraction of organisms killed, and from experimental data and equation (16) we can solve for $g$, which is the expected number of surviving units per organism. Then, from equation (15), when $D$ is large:

$$\log g = \log n - kD. \tag{17}$$

Again extrapolating to $D = 0$ we reach $\log n$. Note that equation (17) has the same form as equation (9), and that the extrapolated line in both cases represents the theoretical survival curve of the units as a function of dose. But $\log g$ will have a linear relation to dose wherever approximation (14) holds, and expression (14) states that the surviving units have a Poisson distribution among the organisms. Thus, if the initial distribution of $n$ (where $D = 0$) resembles more or less a Poisson distribution, then $\log g$ versus $D$ will continue to be more or less linear into the low dose range. If, at $D = 0$, $n$ is a constant as in equation (7), $\log g$ will have spurious values at low doses (Fig. 2, top curve). Intermediate initial distributions will give values of $g$ between these two extremes.

Now suppose that there are two classes of sensitive units with $l$ of one sort and $n$ of another within the same organism, and that the inactivation of all of either class kills the organism. Let $P(A)$ be the probability that an organism survives process A affecting the class of units of number $l$ and $P(B)$ be the probability that an organism survives process B affecting the class of units of number $n$. Then the probability that an organism survives is the probability that it survives both processes,

$$P(AB) = P(A)P(B), \tag{18}$$

if the two processes are independent. From equation (16):

$$P(A) = 1 - e^{-g_1} = 1 - e^{-le^{-k_1D}} \tag{19}$$

and,

$$P(B) = 1 - e^{-g_2} = 1 - e^{-ne^{-k_2D}}.$$

Thus,

$$P(AB) = (1 - e^{-g_1})(1 - e^{-g_2}) = 1 - e^{-g_1} - e^{-g_2} + e^{-(g_1 + g_2)} \tag{20}$$

If $n \gg l$, then at low dose $P(B) \cong 1$ and $P(AB) \cong P(A)$, and $P(A)$ will approximate $1 - e^{-g_1}$ more or less well, depending on the initial distribution of the units of average number $l$, and on the dose reached before $P(B)$ becomes noticeably less than 1. When a sufficiently high dose is reached, expression (20) holds.

Where $S = P(AB)$ is the survival of both processes, and $g \equiv -\ln(1 - S)$, the plot of log $g$ versus $D$ forms two lines (Fig. 3, middle curve). The first line represents the process prevailing in the low dose region, approaches a slope of $-k_1$, and if this slope is reached, extrapolates to $l$ at $D = 0$. The second line represents the combined processes, approaches a slope of $-(k_1 + k_2)$ and extrapolates to $ln$ at $D = 0$. The extension to more than two processes is obvious. It is also evident that in practice, the possibility of inferring two or more processes is dependent on the relative values of $n$, $l$, $k_1$, $k_2$, etc. The presence of an inflection point in the $g$ curve (as in Fig. 3, lower curve) will, however, in general indicate more than one process even though the process affecting the class of units of lowest number may not have proceeded far enough to validate equation (17) before it is masked by the next process.

Let us now consider the multi-process case likely to be encountered if the effect of radiation is predominantly on the genetic apparatus of organisms in which the genetic material is present in more than one set. Let $n$ be the number of sets of genes, each set consisting of $m$ genes, or each set being of such length as to include $m$ deficiencies of average length. From equations (18) and (7):

$$S = \prod_{i=1}^{m} [1 - (1 - e^{-kD})^n] \tag{21}$$

In particular, where $k_1 = k_2 = \ldots = k_m = k$, i.e., where all genes or regions have the same sensitivity:

$$S = [1 - (1 - e^{-k_iD})^n]^m \tag{22}$$

At high dose, by the usual approximation,

$$S \cong n^m e^{-mkD} \tag{23}$$

and extrapolation to $D = 0$ of the linear portion of log $S$, whose slope is $-mk$, gives the value log $(n)^m$.

The above may be summarized as follows:

($a$) For curves of the form of equations (7) or (10) a semi-log plot of $S$ versus $D$ is a curve approaching a line of slope $-k$ as $D$ increases.

($b$) Extrapolation of the linear portion of such a curve to $D = 0$ gives $n$.

($c$) If at $D = 0$, $n$ has a distribution, then extrapolation as above gives the average value of $n$.

($d$) The shape of the curve is altered by such a distribution and also by the eventuality that killing by two or more independent processes occurs.

($e$) A multi-process effect changes the slope of the linear portion of the curve and changes the meaning of its intercept at $D = 0$. The presence of these complications is more evident on inspection if $g \equiv -\ln(1 - S)$ is plotted instead of $S$.[11]

*Experimental.*—To show how the theory yields a reasonable result when applied to experimental data and how the $g$ plot facilitates interpretation some data on ultra-violet irradiation of *Neurospora crassa* macroconidia will be presented (to be reported in greater detail elsewhere). Since the macroconidia are known to be multi-nucleate, equation (21) will hold at least with respect to processes occurring within the nuclei.

The ultra-violet source was a 30-inch Westinghouse sterilamp low-pressure mercury vapor tube stated to deliver 85 per cent of its energy in the 2537 Å. region. Suspensions containing ca. 200,000 conidia per

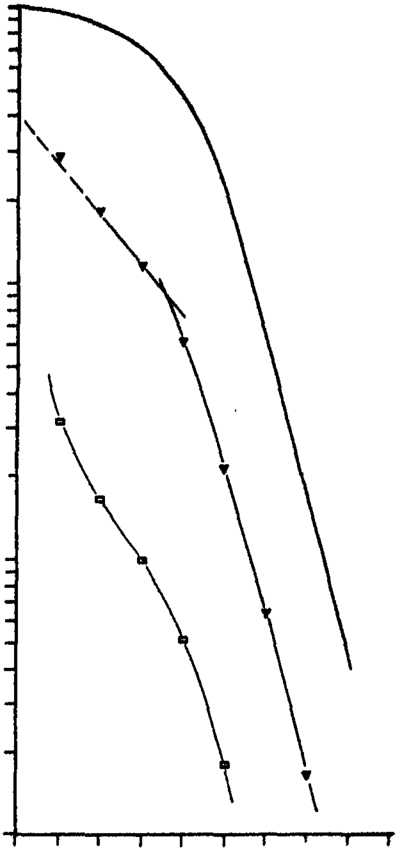


FIGURE 3

Ordinate, $S$ for upper curve, $g$ for lower curves. Middle curve has been moved down one decade and lowest curve two decade. Abscissa, dose in arbitrary units. Upper curve: Theoretical curve for simultaneous survival of two processes, one with $k = 0.43$ and the distribution of $n$ given in Fig. 2, and the other having $k = 1$ and $n = 100$. Middle curve: $g$ computed from upper curve. Lower curve: $g$ computed from theoretical survival of two processes, one having $k = 0.43$ and $n = 3$ (constant), and the other having $k = 1$ and $n = 100$.

cubic centimeter in 0.9 per cent saline were exposed in an open petri plate at a distance of 15 cm. from the center of the tube. The plate was tilted at an angle of 15° from the horizontal and slowly rotated throughout exposure. The average depth of suspension was 5 mm.

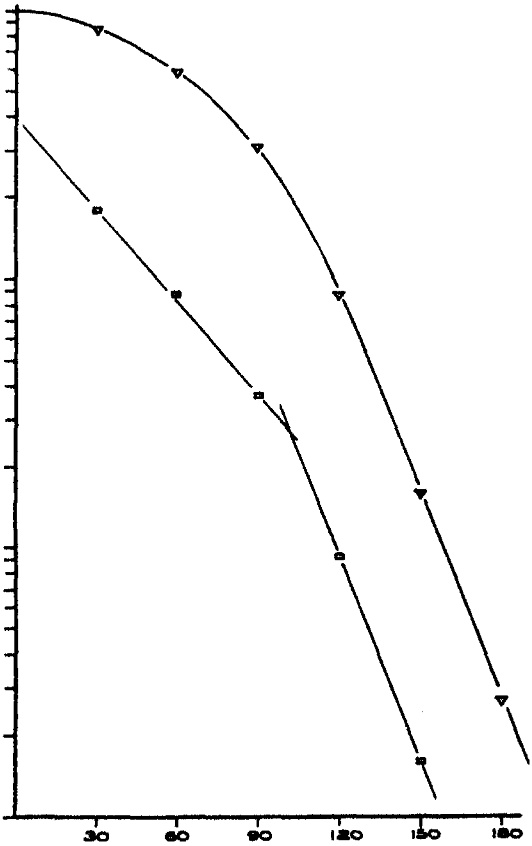


FIGURE 4

Ordinate, *S* for upper curve, *g* for lower curve. Lower curve moved down one decade. Abscissa, seconds of exposure to ultra-violet light under conditions given in text. Upper curve: Survival of macroconidia of *Neurospora crassa* heterokaryon (1633-422)A-(5531-4637)A. Lower curve: *g*, computed from experimental points.

The strain used was a heterokaryon having as one component the double mutant aminobenzoicless-morphological, (1633-422)A, and the other the double mutant pantothenicless-albino (5531-4637)A. This gives normal growth and produces macroconidia on minimal medium. The criterion of viability was the ability to form a visible colony in a sorbose medium.[12] Plates were incubated at 37° in darkness for five days. All handling of irradiated suspensions was done by the light of red photographic safe-lights to prevent photoreactivation.[13]

Table 1 shows the survival relative to exposure time in seconds and the corresponding values of $g$. Figure 4 shows the semi-log plot of $S$ and $g$.

We note from the $g$ plot in figure 4 that the conidia seem to be killed by more than one process, the low dose portion having an intercept at 4. Subsequently, the intercept assumes a minimum value of ca. 130. Further, since the values of $g$ form nearly a straight line at low dose, we conclude that 4 is an average value for the number of some sensitive unit present in the conidia at $D = 0$.

Figure 5 shows the distribution of nuclei in the conidia of the stock used in this experiment, as determined by counts done on preparations fixed and stained by a modification of Robinow's technique.[14] The average is 2.67 nuclei per conidium. If the killing at low dose is related to the number of nuclei present, then 1.5 hits per nucleus are required to agree with $n = 4$. Independent findings of Goodgal[15] and Giles[16] using micro-conidia known to be uninucleate show that these have a 1.5 hit curve with ultra-violet. Thus, it seems probable that the 4-hit component found here is related to the number of nuclei in the macroconidia.

TABLE 1

THE SURVIVAL OF MACROCONIDIA EXPOSED TO ULTRA-VIOLET LIGHT

| Time, seconds | $S$ | $1-S$ | $g$ |
|---|---|---|---|
| 30 | 0.84 | 0.16 | 1.8 |
| 60 | 0.59 | 0.41 | 0.89 |
| 90 | 0.31 | 0.69 | 0.37 |
| 120 | 0.088 | 0.912 | 0.092 |
| 150 | 0.016 | 0.984 | 0.016 |
| 180 | 0.0027 | 0.9973 | 0.0027 |

This relationship must mean that there is some region in each nucleus which is especially sensitive and is responsible for practically all the effect at low dose. If the entire curve is interpreted according to equation (21) we would conclude that at higher doses additional less sensitive regions begin to contribute to the effect. An alternative interpretation based on equation (20) is that there is a very sensitive region as above, but that there is another single independent process involving ca. 30 units of greater sensitivity, and not necessarily related to the nuclei. Obviously, further criteria are necessary to distinguish between these possibilities. The reason for the finding of a 1.5-hit curve for uninucleate conidia is not yet clear.

*Discussion.*—The reasons for preferring the single-unit action interpretation of exponential curves have been discussed by Lea, Haines and

Coulson.[3] Since the theory presented here is based on the assumption that organisms contain units which themselves are inactivated exponentially, the arguments advanced by Lea, *et al.*, in favor of the single-unit action interpretation are again applicable. If we accept these, then the question of whether the theory is reasonable will depend on our reasons for believing that microorganisms showing multi-hit survival contain a number of discrete units rather than a target requiring a certain number of hits.

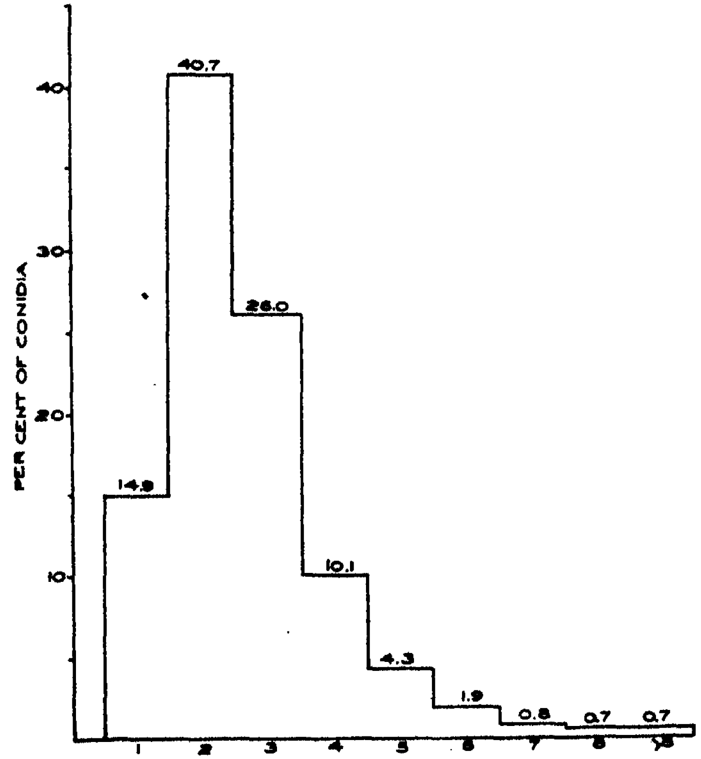


FIGURE 5

Distribution of nuclei in conidia of stock used in Fig. 4; 2134 conidia were scored.

In the special cases we have found where equation (7) has been used previously, the reasons for preferring it to equation (2) are obvious: One of these cases is an explanation of spurious multi-hit curves due to clumping of bacteria[3] and the other is the experiment of Luria and Latarjet designed to follow the increase in numbers of intracellular bacteriophage by means of the survival curves of the infective centers.[7] Equation (22) has recently been given by Luria and Dulbecco[17] in the estimation of the number of genetic units in a phage particle. In other cases where values of $n$ are reported (see review by Zimmer[18]), they have apparently been obtained by fitting the experimental data to theoretical curves prepared from equation (2).

A strong argument for the more general use of equation (21) as a working hypothesis is the repeated demonstration in various microorganisms of genetic mechanisms comparable to those in higher organisms. It is evident that organisms in which the genetic material is duplicated in some way, as by diploidy, polyploidy or multi-nucleatedness, may be expected to give multi-hit curves, and these should be of type (21), but not type (2), in so far as the effects of radiation on the genetic apparatus affect viability. In general, where lethality is due to the exhaustive inactivation of groups of identical essential parts, equation (2) must be rejected.

In the case of a large number, $m$, of identical units any $n \ll m$ of which are indispensible, the probability of obtaining an effective hit would not be decreased significantly even when $n - 1$ of the units had already been altered. This satisfies the requirements of equation (2). However, it seems unlikely that an organism could be killed by a process which is so far from exhaustive.

There are further considerations which make equation (2) seem untenable even when biological criteria are lacking. This becomes clear when we try to think of the observed biological effects in terms of underlying molecular transformations. Assume that a dose, $D$, has a probability, $p$, of changing one of $n$ molecules. Subsequently, an identical dose will have a probability less than $p$ of causing the same change in one of the $n - 1$ remaining unaffected molecules. This is true regardless of whether the transformation of the molecule results from the absorption of a quantum directly (since the probability of such an absorption is proportional to the number of molecules present) or whether the transformation is the end result of a series of chemical reactions, since the probability of occurrence of the reaction which involves the essential molecule is also proportional to the number of such molecules present.

This situation is contrary to the assumption underlying equation (2). In attempting to construct models consistent with the assumptions underlying equation (2) we encounter serious difficulties. For example, a model based on a molecule requiring the absorption of $n$ quanta before it is transformed is objectionable because it implies an unheard of stability of the intermediate excited states, and also leads to intensity dependence. In short, the existence of such a molecule seems contrary to established photochemical principles.

It should be emphasized that equation (7) does not imply any necessity for regarding the units as targets upon which a hit must be scored directly, since the units themselves would show exponential survival regardless of whether the action is direct or indirect.[5] Thus, with radiations giving different ion-densities any differences in the survival curves which are found to be inconsistent with target geometry do not invalidate equation (7) or its corollaries as a model for multi-hit curves. The finding of lowered values of $n$ with radiations of high ion density[19] is consistent with the theory whether or not there is significant indirect effect. With direct effect, this would be due to the increased probability of an ionizing particle traversing more than one unit, and since ionizations produced between the units would be ineffective, there would be a concomitant decrease in $k$. With indirect effect, $n$ would also decrease with increasing ion density if each ionizing particle produced sufficient ionizations to inactivate more than one unit, but here no change in $k$ would be expected since each ionization would have an equal chance of being effective. In situations described by equation (21) an increase in $m$ might be found with high energy radiations due to the equalizing of sensitivity between the various regions or genes involved. As an isolated effect this would increase the over-all value of $k$. However, as we can hardly expect any of these effects to be observed in pure form, it seems most reasonable in comparing multi-hit curves with different types of radiation to consider the differences in terms of a balance between the various factors tending to shift the parameters $n$, $m$ and $k$.

*Summary.*—It is proposed that sigmoid survival curves can be interpreted profitably on a multi-unit, single hit per unit hypothesis, and some consequences of this hypothesis are examined. An aid to the analysis of experimental data is given and its use is illustrated. Reasons are given for preferring the multi-unit hypothesis to the usual multi-hit hypothesis. Some data on the ultra-violet irradiation of *Neurospora crassa* conidia are analyzed.

---

* This work was supported in part by grants from the American Cancer Society, recommended by the Committee on Growth of the National Research Council, and from the Division of Research Grants and Fellowships of the National Institutes of Health, U. S. Public Health Service, both administered by Prof. Francis J. Ryan of Columbia University. The authors are grateful to Professor Ryan for helpful criticism.

† A. E. C. pre-doctoral fellow in biophysics.

[1] Wyckoff, R. W. G., *J. Gen. Physiol.*, **15**, 351–361 (1932).

[2] Lea, D. E., Haines, R. B., and Coulson, C. A., *Proc. Roy. Soc. (Lond.)*, *B*, **120**, 47–76 (1936).

[3] Gates, F. L., *J. Gen. Physiol.*, **13**, 231–248 (1929).

[4] Pugsley, A. T., Oddie, T. H., and Eddy, C. E., *Proc. Roy. Soc. (Lond.)*, *B*, **118**, 276–298 (1935).

[5] Lea, D. E., *Actions of Radiations on Living Cells*, xii + 402 pp., Cambridge Univ. Press (1947).

[6] Latarjet, R., *Ann. Inst. Pasteur*, **70**, 277–285 (1944).

[7] Luria, S. E., and Latarjet, R., *J. Bact.*, **53**, 149–163 (1947).

[8] Sommermeyer, K., *Fiat Rev. Germ. Sci.*, *Biophysics*, Part 1, 15 (1948).

[9] Zimmer, K. G., *Biol. Zentral*, **61**, 208–220 (1941).

[10] Zuppinger, A., *Strahlentherapie*, **28**, 639–758 (1928).

[11] Set $e^{-g} = 1 - S$ where $1 - S$, the fraction killed, is taken from data, compute $g$, then plot log $g$ versus $D$.

[12] Tatum, E. L., R. W. Barratt and V. M. Cutter, Jr., *Science*, **109**, 509–511 (1949). The medium used here contained 1 per cent sorbose, 0.1 per cent sucrose, Fries salts, 4 micrograms biotin per liter, 2 mg. calcium pantothenate per liter, and 2 per cent washed agar.

[13] Kelner, A., Proc. Natl. Acad. Sci., **35**, 73–79 (1949).

[14] Robinow, C. F., *Proc. Roy. Soc. (Lond.)*, *B*, **130**, 299–324 (1941).

[15] Goodgal, S., unpublished (1949).

[16] Giles, N., unpublished (1949).

[17] Luria, S. E., and Dulbecco, R., *Genetics*, **34**, 93–125 (1949).

[18] Zimmer, K. G., *Biol. Zentral.*, **63**, 72–107 (1943).

[19] Zirkle, R., unpublished (1949).

[20] Natural logarithms have been converted to $\log_{10}$ except where the $D = 0$ intercept of an extrapolated line would be changed by so doing.

---

# *ATYPICAL PIGMENT-CELL DIFFERENTIATION IN EMBRYONIC TELEOSTEAN GRAFTS AND ISOLATES*

By Jane M. Oppenheimer

Department of Biology, Bryn Mawr College, and Osborn Zoological Laboratory, Yale University

Communicated by J. S. Nicholas, October 10, 1949

The occasional occurrence of red blood corpuscles and of chromatophores in otherwise non-differentiating isolates and grafts from teleostean gastrulae has been casually mentioned in an earlier publication[1] concerned with other problems than pigment-cell formation. In view of the increasing interest in pigment-cell formation since that time, ensuing upon DuShane's[2] establishment of the neural crest origin of the melanophores in the amphibian, it now seems relevant to present some of the data concerning the atypical differentiation of teleostean chromatophores.

*Fundulus Germ-Ring Grafts.*—In this series of experiments, portions of the germ-ring located 90° or 180° from the midline of the embryonic shield of *Fundulus heteroclitus* gastrulae (cf. Fig. 1 *B*, Oppenheimer[1]) were grafted to the embryonic shield or to the extra-embryonic membrane of gastrulae of the same species. In thirty-seven cases either the whole grafts or portions of them continued development but failed to undergo typical histogenesis and differentiated no axial structures. In sixteen of these, however, red blood corpuscles differentiated, and in ten of these sixteen grafts melanophores differentiated. The ten grafts differentiating melanophores included five implanted on yolk-sac epithelium, and three which were implanted onto the embryonic shield but which, in whole or in part, remained external to the host epithelium and failed to differentiate typically. In the remaining two cases parts of the grafts were incorporated into the host embryos and differentiated axial structures, but adjacent portions of them remained undifferentiated.

In some cases, only a single melanophore was differentiated within the grafts, in others a larger number was present. Only those grafts are included among the ten positive cases enumerated above which contained melanophores located more or less centrally in the mass of cells; grafts to which melanophores were applied only at the surface are excluded from consideration in the possibility that in these cases host melanophores might have migrated to cover the grafts.

In the absence of heteroplastic grafting, there is indirect evidence only that the melanophores were derived from graft rather than host cells. In several cases the melanophores were considerably more densely packed in the grafts than in the adjacent host tissues, and there is no particular reason to suppose that the grafts should attract chromatophores from the host tissues in such numbers as to trap them in higher concentration than usual. In some cases unbroken host epithelium separated graft and host tissues, presumably presenting a barrier to migration. The graft melanophores were occasionally intermingled with the masses of red blood corpuscles which are known to have developed *in situ* since connecting blood vessels between graft and host were absent.

Only heteroplastic transplantation or the grafting of materials from vitally stained embryos can furnish direct proof that the melanophores are formed by graft rather than host cells in such experiments as these. Fortunately, however, supplementary data are available from other types of experiment which demonstrate that chromatophores are also similarly differentiated in some cases by isolates rather than transplants from teleostean gastrulae.

*Epiplatys Germ-Ring Isolates.*—The embryos of *Epiplatys fasciolatus* can be divided at late gastrula stages in such a way (cf. Fig. 1 *A*, Oppenheimer[1]) that the portion of the germ-ring most remote from the embryonic shield can be isolated with part of the yolk from the portion of the egg containing the shield. Such isolates develop for several days in Ringer's solution. Though in some cases they exhibit certain phenomena of growth which simulate the form of the tail, they undergo no histogenesis of axial structures. Of twelve such aggregates studied, one developed red blood corpuscles, one formed a single melanophore and a third differentiated a number of xanthophores. One explant of prospective tailbud region, isolated immediately after the closure of the blastopore, formed both melanophores and xanthophores although otherwise undifferentiated.

*Fundulus Isolated Blastoderms.*—Blastoderms of *Fundulus heteroclitus* separated from the yolk and cultivated in double-strength Holtfreter's solution under some conditions form hyperblastulae, masses of non-differentiated cells provided with a large vesicle at one pole.[3] Of approximately seventy-five hyperblastulae studied, one only developed a single melanophore of typical size and configuration.

*Discussion.*—The production of chromatophores by the grafts and isolates described above suggests that in the teleost under certain experimental conditions pigment-cells can be differentiated by cells which normally do not contribute to the teleostean counterpart of the neural crest. The 90° germ-ring of *Fundulus* normally contributes primarily mesoderm to the embryo;[4] and the 180° germ-ring is unrelated to nervous system formation except in so far as it contributes to the tailbud blastema from which neural derivatives later may arise. While it is theoretically possible, it is highly improbable that the chromatophores in these cases have been differentiated *only* by the particular cells destined to form them later after their passage through the tailbud blastema. The differentiation of pigment-cells under these conditions is therefore presumably the result of what the experimental embryologists call "bedeutungsfremde Selbstdifferenzierung."

This interpretation, however, is not incompatible with the possibility that in teleosts the chromatophores normally arise from cells corresponding to those of the neural crest of other vertebrates, as suggested by the results of the transplantation experiments of Lopashov[5] on three species of teleosts, and in line with the neural crest origin of pigment-cells in amphibians,[2] birds[6] and mammals.[7] It merely signifies that under certain abnormal conditions, other than the usual cells can take over the function of pigment-cell formation.

The differentiation of red blood corpuscles in the same series of preparations, and in some cases by the same grafts that differentiated the chromatophores, suggests that the production of these cell types is the result of a generalized type of reaction. Somewhat comparable differentiation of red blood corpuscles in otherwise non-differentiating grafts has been described for the chick;[8] and somewhat analogous to this condition may be the differentiation in chick embryonic grafts of ganglion cells formed in the absence of differentiation of central nervous tissue.[9]

Whatever the factors may be in these grafts and isolates that suppress their typical differentiation or that fail to support it, they seem to be allied with factors that foster the atypical production of chromatophores by cells originally destined to fulfil other functions. The fact that such conditions obtain is of twofold significance. In the first place, they suggest that in the teleosts at least, great caution must be exercised in the interpretation of transplantation- or explantation-experiments designed to elucidate the normal development of pigment-cells. More positively, they offer the hope that since in these forms pigment-cells can be produced under experimental conditions by other than typically differentiating cells, it may prove experimentally feasible in them to analyze some of the factors responsible for this abnormal differentiation, and hence to shed further light on the normal mechanism of production of these cells in the typically differentiating embryo.

*Summary.*—Grafts of germ-ring of *Fundulus heteroclitus* gastrulae, exhibiting no other typical histogenesis, have in some cases been found to differentiate red blood corpuscles and melanophores. Red blood cells, melanophores and occasionally xanthophores have been found present in tail-shaped structures lacking other histological differentiation formed by portions of *Epiplatys fasciolatus* germ ring isolated in Ringer's solution. In a single case out of approximately seventy-five studied, a melanophore was differentiated in the absence of other histogenesis by a hyperblastula formed by an early *Fundulus heteroclitus* blastoderm removed from the yolk and cultivated in double-strength Holtfreter's solution. These facts in no way negate the possibility that in teleosts chromatophores are normally provided by cells comparable to the neural crest of amphibians, birds and mammals. They suggest, however, that under certain experimental conditions not adequate to foster normal histogenesis, cells other than those destined for such differentiation are capable of forming pigment-cells.

[1] Oppenheimer, J. M., *J. Exp. Zool.*, **79**, 185–212 (1938).

[2] DuShane, G. P., *Ibid.*, **72**, 1–31 (1935).

[3] Oppenheimer, J. M., *Ibid.*, **72**, 247–269 (1936).

[4] Oppenheimer, J. M., *Ibid.*, **73**, 405–444 (1936).

[5] Lopashov, G. V., *C. R. (Doklady) Acad. Sci. U.S.S.R.*, **44**, 169–172 (1944).

[6] Dorris, F., *Arch. Entw.-mech.*, **138**, 323–334 (1938).

[7] Rawles, M. E., *Physiol. Zool.*, **20**, 248–266 (1947).

[8] Rudnick, D., *J. Exp. Zool.*, **79**, 399–427 (1938).

[9] Rudnick, D., *Ibid.*, **78**, 369–383 (1938).

---

## ERRATA

In the article "Studies on the Biochemical Genetics of Yeast," these PROCEEDINGS, pages 456–464, August, 1949, the reference[21] on p. 459, 8th line up, and on p. 460, 5th line down, should be [20a], and among the references on p. 464 should be inserted:

[20a] Burkholder, P. R., and Giles, N. H., Jr., *Am. J. Botany*, **34**, 345–348 (1947).

SEYMOUR POMPER

# INDEX

Aberrant heterozygotes (Lederberg), 178.
Action of peptidases (Smith), 80.
ADDIS, T. The Mechanism of Proteinuria, 194.
Adsorbents, specific (Dickey), 227.
Age of members of National Academy of Sciences (Wright), 117.
Aggregation phenomena (Bier and Nord) 17; (Timasheff, Bier and Nord), 364, 494.
Aging (Sherman and Trupp), 90.
ALBERT, A. A. A Theory of Trace-Admissible Algebras, 317.
Alcohol-acetate solutions (Timasheff, Bier and Nord), 364.
Alcoholism (Williams, Berry and Beerstecher), 265.
Algebraic functions (Bellman and Straus), 472.
Algebraic systems (Smith), 405.
Algebras (Albert), 317.
Algebras, group (Kaplansky), 133.
*Alopex-Vulpes* (Wipf and Shackelford), 468.
ANKENY, N. C., and CHOWLA, S. The Class Number of the Cyclotomic Field, 529.
ANTHONY, D. S. See Foster, J. W., 663.
Antibiotic substances (Kavanagh, Hervey and Robbins), 343.
ATWOOD, K. C., and NORMAN, A. On the Interpretation of Multi-Hit Survival Curves, 696.
Automorphisms (Hua), 386.
Axioms for sets (Wang), 150.
Axolotl (Dalton), 277.

Bacteria, mutants (Davis), 1.
Bacteria, ultra-violet inactivation (Novick and Szilard), 591.
BAL (Thimann and Bonner), 272.
Basidiomycetes (Kavanagh, Hervey and Robbins), 343.
Bayes solutions (Wald and Wolfowitz), 99.
Beaded Minute combination (Sturtevant), 311; (Goldschmidt), 314.
BECK, L. H. See Miles, W. R., 292.
BEERSTECHER, E., JR. See Williams, R. J., 265.
BELLMAN, R., and STRAUS, E. G. Continued Fractions, Algebraic Functions and the Padé Table, 472.
BELLMAN, R., and BLACKWELL, D. Some Two-Person Games Involving Bluffing, 600.
Bending of tubes (Reissner), 204.
Bernoulli members (Vandiver), 332.
BERRY, L. J. See Williams, R. J., 265.
BIER, M., and NORD, F. F. Aggregation Phenomena in Egg Albumin Solutions as Determined by Light Scattering Measurements, 17.
BIER, M. See Timasheff, S. N., 364, 494.
Bilinear functionals (Morse and Transue), 136.
Biochemical genetics (Pomper and Burkholder), 456.
Bioelectric potentials (MacInnes), 457; (Osterhout), 548; (Cole), 558; (Blinks) 566.
Birational transformations (Zariski), 62.
BLACKWELL, D. See Bellman, R., 600.
BLAKERS, A. L., and MASSEY, W. S. The Homotopy Groups of a Triad, 322.
BLINKS, L. R. The Source of the Bioelectric Potentials in Large Plant Cells, 566.
Bluffing, games (Bellman and Blackwell), 600.
BOCHNER, S. Diffusion Equation and Stochastic Processes, 368.
Bohr magneton (Condon), 488.
BONNER, D. M., and YANOFSKY, C. Quinolinic Acid Accumulation in the Conversion of 3-Hydroxyanthranilic Acid to Niacin in Neurospora, 576.
BONNER, W. D. JR. See Thimann, K. V., 272.
Bounds for solution of differential equation (Leighton), 190.
Brunn-Minkowski convex bodies (Busemann), 27.
BUERGER, M. J. Crystallographic Symmetry in Reciprocal Space, 198.
BURKHOLDER, P. R. See Pomper, S. 456, 712.
BURNHAM, C. R. Chromosome Segregation in Maize Translocations in Relation to Crossing Over in Interstitial Segments, 349.
BUSEMANN, H. A Theorem on Convex Bodies of the Brunn-Minkowski Type, 27.

$C_2$ condensation (Foster, Carson, Anthony, Davis, Jefferson and Long), 663.
Carbohydrate fermentation (Lindegren and Lindegren), 23.
Carbon monoxide (Pauling and Sheehan), 359.
CARMICHAEL, L., KENNEDY, J. L., and MEAD, L. C. Some Recent Approaches to the Experimental Study of Human Fatigue, 691.
Carotenoid differences (Mackinney and Jenkins), 284.
CARSON, S. F. See Foster, J. W., 663.
CASTLE, W. E., and KING, H. D. Linkage Studies of the Rat. X, 545.
Cauchy-Frullani integral (Ostrowski), 612
Cells, bioelectric potentials (Osterhout), 548; (Cole), 558; (Blinks), 566.
Cells, epidermal (Lewis), 506.

"Central" probability problem (Loève), 328.
Centromere organization (Rhoades and Kerr), 129.
Chaney, R. W. The Miocene Occurrence of Sequoia and Related Conifers in the John Day Basin, 125.
——. Early Tertiary Ecotones in Western North America, 356.
Characters over rings (Hua and Vandiver) 94.
Chen, T.-T. Conjugation Between Double Monsters and Single Animals in *Paramecium Bursaria*, 108.
Chlorophyceae (Starr), 453.
Chloroplast pigment (Everett), 628.
Chowla, S. On Difference Sets, 92.
——. The Last Entry in Gauss's Diary, 244.
Chowla, S., and Selberg, A. On Epstein's Zeta Function (I), 371.
Chowla, S. See Ankeny, N. C., 529.
Chromatid aberrations (Swanson), 237.
Chromatin and methyl green (Pollister and Leuchtenberger), 111.
Chromosomal rearrangements (Giles and Riley), 640.
Chromosome segregation (Burnham), 349.
Chromosomes of fox hybrid (Wipf and Shackelford), 468.
Chung, K. L., and Feller, W. On Fluctuations in Coin-Tossing, 605.
Class number (Ankeny and Chowla), 529.
Cleland, G. H. See Dickey, F. H., 581.
Coefficient problem (Royden), 657.
Coefficients of Schlicht functions (Schaeffer and Spencer), 143.
Coin-tossing (Chung and Feller), 605.
Cole, K. S. Some Physical Aspects of Bioelectric Phenomena, 558.
Condon, E. U. Superconductivity and the Bohr Magneton, 488.
Conference on bioelectric potentials, (MacInnes), 547.
Congruences (Vandiver), 332.
Conjugation in Paramecium (Chen), 108.
Continued fractions (Bellman and Straus) 472.
Continuity, absolute (Rado and Reichelderfer), 678.
Convergence of harmonic polynomials (Walsh, Sewell and Elliott), 59.
Convex bodies (Busemann), 27.
Coumarin (Thimann and Bonner), 272.
Crossing over (Laughnan), 167; (Burnham), 349; (Green and Green), 556.
Crustacean retina (Kleinholz), 215.
Crystallographic symmetry (Buerger), 198.
Culture of *Paramecium aurelia* (Van Wagtendonk and Hackett), 155.
Curved profiles, pressure on (Thomas), 617.
Curves, physical (Kasner and De Cicco), 106.
Cyclotomic field (Ankeny and Chowla), 529.
Cytoplasmic factor in sterility (Everett), 628.

Dalton, H. C. Developmental Analysis of Genetic Differences in Pigmentation in the Axolotl, 277.
Davis, B. D. The Isolation of Biochemically Deficient Mutants of Bacteria by Means of Penicillin, 1.
Davis, J. B. See Foster, J. W., 663.
DeCicco, J. See Kasner, E., 43, 106, 201, 419.
Decision problems (Wald and Wolfowitz), 99.
Desoxyribose nucleic acid (Schrader and Leuchtenberger), 464.
Developmental analysis (Dalton), 277.
Dickey, F. H. The Preparation of Specific Adsorbents, 227.
Dickey, F. H., Cleland, G. H., and Lotz, C. The Rôle of Organic Peroxides in the Induction of Mutations 581.
Dieudonné (Halmos), 38.
Difference sets (Chowla), 92.
Diffusion equation (Bochner), 368.
Differential equations (Kasner and De Cicco), 43; (Kasner and Mittleman), 338; (Leighton), 190, 192, 422, 656.
Diploid *Zea Mays* (Everett), 628.
Dissociation energy (Pauling and Sheehan), 359.
Distance, perception of (Ratoosh), 257.
Distributions (Kac), 252.
*Drosophila melanogaster* (Sturtevant), 311; (Goldschmidt), 314; (Green and Green), 556; (Kerschner); 647.
Dvoretzky, A. On the Theorem of Jentzsch, 246.

Ecotones, Tertiary (Chaney), 356.
Egg albumin (Bier and Nord), 17.
Eigenvalues (Weyl), 408; (Fan), 652.
Eisenhart, L. P. Separation of the Variables in the One-Particle Schroedinger Equation in 3-Space, 412.
——. Separation of the Variables of the Two-Particle Wave Equation, 490.
Elliott, H. M. See Walsh, J. L., 59.
Embryonic teleostean grafts (Oppenheimer), 709.
Energy, valence-state (Pauling), 229.
Entire functions (Rådström), 399.
Epiboly (Trinkaus), 218.
Epidermal cells (Lewis), 506.
Epstein's Zeta function (Chowla and Selberg), 371.
Equations in finite field (Hua and Vandiver), 94, 477, 481.

Erdös, P. On a New Method in Elementary Number Theory Which Leads to an Elementary Proof of the Prime Number Theorem, 374.
Errata (Neményi and Prim), 116; (Leighton), 422; (Timasheff, Bier and Nord), 494; (Pomper and Burkholder), 712.
Erythrocyte, bovine (Stormont), 232.
*Escherichia coli* (Lederberg), 178.
Everett, H. L. A Genic Series Controlling Chloroplast Pigments' Production in Diploid *Zea Mays* L, 628.
Evolution of stars (Struve), 161.
Exponential sums (Vandiver), 686.

Fan, K. On a Theorem of Weyl Concerning Eigenvalues of Linear Transformations I, 652.
Fatigue, human (Carmichael, Kennedy and Mead), 691.
Feller, W. See Chung, K. L., 605.
Fermat quotients (Vandiver), 332.
Fields of force (Kasner and De Cicco), 419.
Finite field (Hua and Vandiver), 477, 481; (Vandiver), 681.
Fluctuations in coin-tossing (Chung and Feller), 605.
Foster, A. L. The *n*-ality Theory of Rings, 31.
Foster, J. W., Carson, S. F., Anthony, D. S., Davis, J. B., Jefferson, W. E., and Long, M. V. Aerobic Formation of Fumaric Acid in the Mold *Rhisopus Nigricans:* Synthesis by Direct $C_2$ Condensation, 663.
Fourier series (Morse and Transue), 395.
Fox hybrid (Wipf and Shackelford), 468.
Fréchet variation (Morse and Transue), 395.
Fumaric acid (Foster, Carson, Anthony, Davis, Jefferson and Long), 663.
Functionals, bilinear (Morse and Transue), 136.
Functionals, quadratic (Leighton), 192.
*Fundulus* eggs (Trinkaus), 218.

Gabelman, W. H. Reproduction and Distribution of the Cytoplasmic Factor for Male Sterility in Maize, 634.
Galston, A. W. Riboflavin-Sensitized Photoöxidation of Indole-Acetic Acid and Related Compounds, 10.
Games (Bellman and Blackwell), 600.
Gauss's diary (Chowla), 244.
Gene-controlled carbohydrate fermentation (Lindegren and Lindegren), 23.
Genetics of yeast (Pomper and Burkholder), 456, 712.
Geneto-trophic diseases (Williams, Berry and Beerstecher), 265.
Geometry of moment spaces (Karlin and Shapley), 673.
Gilbarg, D. A Generalization of the Schwarz-Christoffel Transformation, 609.
Giles, N. G., Jr., and Riley, H. P. The Effect of Oxygen on the Frequency of X-Ray Induced Chromosomal Rearrangements in Tradescantia Microspores, 640.
Gleason, A. M. On the Structure of Locally Compact Groups, 384.
Goldschmidt, R. B. The Intersexual Males of the Beaded Minute Combination in *Drosophila Melanogaster*, 314.
Graham, J. B., and Graham, R. M. Pharmacological Modification of Resistance to Radiation—A Preliminary Report, 102.
Graham, R. M. See Graham, J. B., 102.
Graphite (Pauling and Sheehan), 359.
Grafts, teleostean (Oppenheimer), 709.
Group algebras (Kaplansky), 133.
Groups, homotopy (Blakers and Massey), 322; (Smith), 405.
Groups, locally compact (Mackey), 537.
Groups, structure of (Gleason), 384.
Growth, plant (Thimann and Bonner), 272.
Green, K. C. See Green, M. M., 586.
Green, M. M., and Green, K. C. Crossing-Over Between Alleles at the Lozenge Locus in *Drosophila Melanogaster*, 586.

Hackett, P. L. See van Wagtendonk, W. J., 155.
Halmos, P. R. On a Theorem of Dieudonné, 38.
Harmonic polynomials and functions (Walsh, Sewell and Elliott), 59.
Haskins, F. A., and Mitchell, H. K. Evidence for a Tryptophane Cycle in Neurospora, 500.
Heat of sublimation (Pauling and Sheehan), 359.
Hervey, A. See Kavanagh, F., 343.
Homotopy groups (Blakers and Massey), 322; (Smith), 405.
Honeybees (Miles and Beck), 292.
Hua, L.-K. On the Automorphisms of a Sfield, 386.
——. Some Properties of a Sfield, 533.
Hua, L. K., and Vandiver, H. S. Characters over Certain Types of Rings with Applications to the Theory of Equations in a Finite Field, 94.
——. On the Number of Solutions of Some Trinomial Equations in a Finite Field, 477.
——. On the Nature of the Solutions of Certain Equations in a Finite Field, 481.
Human fatigue (Carmichael, Kennedy and Mead), 691.

Ideals, primary (Kaplansky), 133.
Imprimitivity for representations (Mackey), 537.
Indeterminates, use of (Vandiver), 686.
Indole-acetic acid (Galston), 10.
Infrared absorption (Miles and Beck), 292.
Infra-red radiation (Swanson), 237.
Inheritance of carotenoid differences (Mackinney and Jenkins), 284.
Inhibition of growth (Thimann and Bonner), 272.
Integral, Cauchy-Frullani (Ostrowski), 612.
Integral curves (Kasner and De Cicco), 43.
Integral representations (Morse and Transue), 136.
Integration (McShane), 46; (Stone), 50.
Interposition (Ratoosh), 257.
Isolates, embryonic (Oppenheimer), 709.
Isolation of bacteria by penicillin (Davis), 1.

J substance (Stormont), 232.
Jacobian, generalized (Rado and Reichelderfer), 678.
Jenkins, J. A. See Mackinney, G., 284.
Jefferson, W. E. See Foster, J. W., 663.
Jentzsch (Dvoretzky), 246.
John Day Basin (Chaney), 125.

Kac, M. On Deviations Between Theoretical and Empirical Distributions, 252.
Kaplansky, I. Primary Ideals in Group Algebras, 133.
Karlin, S., and Shapley, L. S. Geometry of Reduced Moment Spaces, 673.
Kasner, E., and De Cicco, J. Osculating Comics of the Integral Curves of Third Order Differential Equation of the Type (G), 43.
——. Physical Systems of Curves in Space, 106.
——. Physical Curves in Space of *n* Dimensions, 201.
——. Physical Families in Conservative Fields of Force, 419.
Kasner, E., and Mittleman, D. Second Order Differential Equations of Rank 2 in Physics, 338.
Kavanagh, F., Hervey, A., and Robbins, W. J. Antibiotic Substances from Basidiomycetes. IV. *Marasmius Conigenus*, 343.
Kelner, A. Effect of Visible Light on the Recovery of *Streptomycens Griseus* Conidia from Ultraviolet Irradiation Injury, 73.
Kennedy, J. L. See Carmichael, L., 691.
Kerr, W. E. See Rhoades, M. M., 129.
Kerschner, J. The Influence of the Y-Chromosome on X-Ray Induced Mutation Rate in *Drosophila Melanogaster*, 647.
King, H. D. See Castle, W. E., 545.
Kleinholz, L. H. Responses of the Proximal Retinal Pigment of the Isolated Crustacean Eyestalk to Light and to Darkness, 215.
Kosambi, D. D. Lie Rings in Path Space, 389.
Kroeber, A. L. Values as a Subject of Natural Science Inquiry, 261.

Laughnan, J. R. The Action of Allelic Forms of the Gene A in Maize, II. The Relation of Crossing Over to Mutation of $A^b$, 167.
Length of Life (Sherman and Trupp), 90.
Lederberg, J. Aberrant Heterozygotes in *Escherichia Coli*, 178.
Leighton, W. Bounds for the Solutions of a Second-Order Linear Differential Equation, 190, 422.
——. Principal Quadratic Functionals and Self-Adjoint Second-Order Differential Equations, 192.
——. On Self-Adjoint Differential Equations of Second Order, 656.
Leuchtenberger, C. See Pollister, A. W., 66, 111. See Schrader, F., 464.
Lewis, F. T. The Correlation in Shape and Size Between Epidermal and Subepidermal Cells, 506.
Lie rings (Kosambi), 389.
Lifting surfaces (Reissner), 208.
Light-reactivation (Novick and Szilard), 591.
Light, response to (Kleinholz), 215.
Light scattering (Bier and Nord), 17.
Lindegren, C. C., and Lindegren, G. Unusual Gene-Controlled Combinations of Carbohydrate Fermentations in Yeast Hybrids, 23.
Lindegren, G. See Lindegren, C. C., 23.
Linkage in rat (Castle and King), 545.
Loève, M. On the "Central" Probability Problem, 328.
Long, M. V. See Foster, J. W., 663.
Lotz, C. See Dickey, F. H., 581.
Lycopersicon esculentum (MacKinney and Jenkins), 284.

MacInnes, D. A. The Conference on Bioelectric Potentials, 547.
Mackey, G. W. Imprimitivity for Representations of Locally Compact Groups, I, 537.
MacKinney, G., and Jenkins, J. A. Inheritance of Carotenoid Differences

in Lycopersicon Esculentum Strains, 284.
Magneton (Condon), 488.
Maize, gene A (Laughnan), 167.
Maize, sterility in (Gabelman), 634.
Maize translocations (Burnham), 349.
*Marasmius conigenus* (Kavanagh, Hervey and Robbins), 343.
Massey, W. S. See Blakers, A. L., 322.
McElroy, W. D. See Swanson, C. P., 513.
McShane, E. J. Remark Concerning Integration, 46.
Mead, L. C. See Carmichael, L., 691.
Membership of National Academy of Sciences, distribution and ages of (Wright), 117.
Metabolic patterns (Williams, Berry and Beerstecher), 265.
Methyl green and chromatin (Pollister and Leuchtenberger), 111.
Metal-peptidases (Smith), 80.
Microspores, Tradescantia (Giles and Riley), 640.
Miles, W. R., and Beck, L. H. Infrared Absorption in Field Studies of Olfaction in Honeybees, 292.
Miller, H. See Swanson, C. P., 513.
Miocene Sequoia (Chaney), 125.
Mitchell, H. K. See Haskins, F. A., 500.
Mittleman, D. See Kasner, E., 338.
Monsters in Paramecium (Chen), 108.
Morse, M., and Transue, W. Integral Representations of Bilinear Functionals, 136.
——. The Fréchet Variation and the Convergence of Multiple Fourier Series, 395.
Multi-hit survival curves (Atwood and Norman), 696.
Mutants of bacteria (Davis), 1.
Mutation and organic peroxides (Dickey, Cleland and Lotz), 581.
Mutation and Y-chromosome (Kerschner), 647.
Mutation of $A^b$ in maize (Laughnan), 167.
Mutation rate (Swanson, McElroy and Miller), 513.

*n*-ality theory of rings (Foster), 31.
National Academy of Sciences, The Conference on Bioelectric Potentials (MacInnes, Osterhout, Cole and Blinks), 547–575.
National Academy of Sciences, Distribution and Ages of Membership (Wright), 117.
National Academy of Sciences—Organization, 423.
Neményi, P., and Prim, R. See Errata.
*Neurospora* (Wagner), 155; (Haskins and Mitchell), 500.
Neurospora, niacin in (Bonner and Yanofsky), 576.
Niacin in Neurospora (Bonner and Yanofsky), 576.
Nitrogen mustard (Swanson, McElroy and Miller), 513.
Nord, F. F. See Bier, M., 17. See Timasheff, S. N., 364, 494.
Norman, A. See Atwood, K. C., 696.
Novick, A., and Szilard, L. Experiments on Light-Reactivation of Ultra-Violet Inactivated Bacteria, 591.
Nucleic acid (Schrader and Leuchtenberger), 464.
Nucleoprotein content of nuclei (Pollister and Leuchtenberger), 66.
Numbers of solutions of equations (Vandiver), 681.

Ocean wave data (Seiwell), 518.
Olfaction (Miles and Beck), 292.
Oppenheimer, J. M. Atypical Pigment-Cell Differentiation in Embryonic Teleostean Grafts and Isolates, 709.
Organic peroxides in mutation (Dickey, Cleland and Lotz), 581.
Organization, centromere (Rhoades and Kerr), 129.
Organization of National Academy of Sciences, 423.
Osculating conics (Kasner and De Cicco), 43.
Osterhout, W. J. V. Some Bioelectrical Problems, 548.
Ostrowski, A. M. On Some Generalizations of the Cauchy-Frullani Integral, 612.
Oxygen atom (Pauling), 229.
Oxygen, effect on chromosomal rearrangements (Giles and Riley), 640.

Padé table (Bellman and Straus), 472.
Pantothenic acid (Wagner), 185.
*Paramecium aurelia* (van Wagtendonk and Hackett), 155.
*Paramecium bursaria* (Chen), 108.
Path space (Kosambi), 389.
Pauling, L. The Valence-State Energy of the Bivalent Oxygen Atom, 229.
——. On the Stability of the $S_8$ Molecule and the Structure of Fibrous Sulfur, 495.
Pauling, L., and Sheehan, W. F., Jr. The Dissociation Energy of Carbon Monoxide and the Heat of Sublimation of Graphite, 359.
Penicillin (Davis), 1.
Perception of distance (Ratoosh), 257.
Pharmacological modification of resistance (Graham and Graham), 102.
Photoöxidation of indole-acetic acid (Galston), 10.

Physical curves (Kasner and De Cicco), 106, 201, 419.
Physics, differential equations in (Kasner and Mittleman), 338.
Pigment production (Everett), 628.
Pigmentation (Dalton), 277.
Pigment-cell differentiation (Oppenheimer), 709.
Plant growth (Thimann and Bonner), 272.
Pollister, A. W., and Leuchtenberger, C. The Nucleoprotein Content of Whole Nuclei, 66.
——. The Nature of the Specificity of Methyl Green for Chromatin, 111.
Polyvinyl alcohol-acetate solutions (Timasheff, Bier and Nord), 364, 494.
Pomper, S., and Burkholder, P. R. Studies on the Biochemical Genetics of Yeast, 456, 712.
Potentials, bioelectric (MacInnes), 547, (Osterhout), 548; (Cole), 558; (Blinks), 566.
Pressure distribution (Thomas), 617.
Prime number theorem (Erdös), 374.
Probability (Loève), 328.
Proteinuria (Addis), 194.
Protoanemonin (Thimann and Bonner), 272.

Quadratic relations (Vandiver), 681.
Quinolinic acid (Bonner and Yanofsky), 576.

Radiation, resistance to (Graham and Graham), 102.
Rado, T., and Reichelderfer, P. V. On $n$-Dimensional Concepts of Bounded Variation, Absolute Continuity and Generalized Jacobian, 678.
Rådström, H. On the Zero of Successive Derivatives of Entire Functions, 399.
Rat, linkage in (Castle and King), 545.
Ratoosh, P. On Interposition as a Cue for the Perception of Distance, 257.
Reciprocal space (Buerger), 198.
Reduced moment spaces (Karlin and Shapley), 673.
Reichelderfer, P. V. See Rado, T., 678.
Reissner, E. On Bending of Curved Thin-Walled Tubes, 204.
——. Note on the Theory of Lifting Surfaces, 208.
Representations of groups (Mackey), 537.
Resistance to radiation (Graham and Graham), 102.
Response to light (Kleinholz), 215.
Retina, crustacean (Kleinholz), 215.
*Rhizopus nigricans* (Foster, Carson, Anthony, Davis, Jefferson and Long), 663.
Rhodes, M. M., and Kerr, W. E. A Note on Centromere Organization, 129.
Riboflavin-sensitized photoöxidation (Galston), 10.
Riley, H. P. See Giles, N. H., Jr., 640
Rings, characters over (Hua and Vandiver), 94.
Rings, Lie (Kosambi), 389.
Rings, theory of (Foster), 31.
Robbins, W. J. See Kavanagh, F., 348.
Royden, H. L. The Coefficient Problem for Bounded Schlicht Functions, 657.

$S_8$ molecule (Pauling), 495.
Schaeffer, A. C., and Spencer, D. C. The Coefficients of Schlicht Functions, IV, 143.
Schlicht functions (Schaeffer and Spencer), 143; (Royden), 657.
Schrader, F., and Leuchtenberger, C. Variation in the Amount of Desoxyribose Nucleic Acid in Different Tissues of Tradescantia, 464.
Schroedinger equation (Eisenhart), 412.
Schwarz-Christoffel transformation (Gilbarg), 609.
Seiwell, H. R. The Principles of Time Series Analyses Applied to Ocean Wave Data, 518.
Selberg, A. See Chowla, S., 371.
Self-adjoint differential equations (Leighton), 656.
Sequential decision problems (Wald and Wolfowitz), 99.
Sequoia, Miocene (Chaney), 125.
Set theory (Wang), 150.
Sets, difference (Chowla), 92.
Sewell, W. E. See Walsh, J. L., 59.
Sex-determination (Sturtevant), 311; (Goldschmidt), 314.
Sfield (Hua), 386, 533.
Shackelford, R. M. See Wipf, L., 468.
Shape and size of cells (Lewis), 506.
Shapley, L. S. See Karlin, S., 673.
Sheehan, W. F., Jr. See Pauling, L., 359.
Sherman, H. C., and Trupp, H. Y. Further Experiments with Vitamin A in Relation to Aging and to Length of Life, 90.
Smith, E. L. The Mode of Action of the Metal-Peptidases, 80.
Smith, P. A. Homotopy Groups of Certain Algebraic Systems, 405.
Solutions of Equations (Vandiver), 681.
Space, physical curves in (Kasner and De Cicco), 201.
Specific adsorbents (Dickey), 227.
Spencer, D. C. See Schaeffer A. C., 143.
Stability of $S_8$ (Pauling), 495.
Starr, R. C. A Method of Effecting

Zygospore Germination in Certain Chlorophyceae, 453.
Stars, evolution of (Struve), 161.
Sterility in maize (Gobelman), 628.
Stochastic processes (Bochner), 368.
Stone, M. H. Notes on Integration, IV, 50.
Stormont, C. Acquisition of the J Substance by the Bovine Erythrocyte, 232.
Straus, E. G. See Bellman, R., 472.
*Streptomyces griseus* (Kelner), 73.
Structure of groups (Gleason), 384.
Struve, O. Some Problems of the Evolution of Stars, 161.
Sturtevant, A. H. The Beaded Minute Combination and Sex-Determination in Drosophila, 311.
Sulfur, fibrous (Pauling), 495.
Superconductivity (Condon), 488.
Supersonic speeds (Thomas), 617.
Surface gel layers (Trinkaus), 218.
Surfaces, lifting (Reissner), 208.
Survival curves (Atwood and Norman), 696.
Swanson, C. P. Further Studies on the Effect of Infra-Red Radiation on X-Ray-Induced Chromatid Aberrations in *Tradescantia*, 237.
Swanson, C. P., McElroy, W. D., and Miller, H. The Effect of Nitrogen Mustard Pretreatment on the Ultra-Violet-Induced Morphological and Biochemical Mutation Raté, 513.
Symmetry, crystallographic (Buerger), 198.
Szilard, L. See Novick, A., 591.

Tertiary ecotones (Chaney), 356.
Theoretical and empirical distributions (Kac), 252.
Thimann, K. V., and Bonner, W. D., Jr. Inhibition of Plant Growth by Protoanemonin and Coumarin, and Its Prevention by BAL, 272.
Thomas, T. Y. First Approximation of Pressure Distribution on Curved Profiles at Supersonic Speeds, 617.
Timasheff, S. N., Bier, M., and Nord, F. F. Aggregation Phenomena in Polyvinyl Alcohol-Acetate Solutions, 364, 494.
Time series (Seiwell), 518.
Trace-admissible algebras (Albert), 317.
*Tradescantia* (Swanson), 237; (Schrader and Leuchtenberger), 464.
Tradescantia microspores (Giles and Riley), 640.
Transformation, linear (Weyl), 408; (Fan), 652.
Transformation, Schwarz-Christoffel (Gilbarg), 609.
Transformations, birational (Zariski), 62.
Transue, W. See Morse, M., 136, 395.
Trinkaus, J. P. The Surface Gel Layer of Fundulus Eggs in Relation to Epiboly, 218.
Trinomial equations (Hua and Vandiver), 477.
Trupp, H. Y. See Sherman, H. C., 90.
Tryptophane cycle (Haskins and Mitchell), 500.
Tubes, bending of (Reissner), 204.
Two-person games (Bellman and Blackwell), 600.

Ultra-violet inactivation of bacteria (Novick and Szilard), 591.
Ultra-violet-induced mutation (Swanson, McElroy and Miller), 513.
Ultra-violet irradiation injury (Kelner), 73.

Valence-state energy (Pauling), 229.
Values (Kroeber), 261.
Vandiver, H. S. On Congruences Which Relate the Fermat and Wilson Quotients to the Bernoulli Numbers, 332.
——. Quadratic Relations Involving the Numbers of Solutions of Certain Types of Equations in a Finite Field, 681.
——. On the Use of Indeterminates in the Theory of Exponential Sums, 686.
Vandiver, H. S. See Hua, L. K., 94, 477, 481.
van Wagtendonk, W. J., and Hackett, P. L. The Culture of *Paramecium Aurelia* in Absence of Other Living Organisms, 155.
Variation, bounded (Rado and Reichelderfer), 678.
Vitamin A (Sherman and Trupp), 90.
von Neumann's axioms for sets (Wang), 150.

Wagner, R. P. The *in vitro* Synthesis of Pantothenic Acid by Pantothenicless and Wild Type *Neurospora*, 185.
Wald, A., and Wolfowitz, J. Bayes Solutions of Sequential Decision Problems, 99.
Walsh, J. L., Sewell, W. E., and Elliott, H. M. On the Degree of Convergence of Harmonic Polynomials to Harmonic Functions, 59.
Wang, H. On Zermelo's and von Neumann's Axioms for Set Theory, 150.
Wave analyses, ocean (Seiwell), 518.
Wave-equation (Eisenhart), 490.
Weyl, H. Inequalities Between the Two Kinds of Eigenvalues of a Linear Transformation, 408.
Weyl on eigenvalues (Fan), 652.
Williams, R. J., Berry, L. J., and Beerstecher, E., Jr. Individual Meta-

bolic Patterns, Alcoholism, Genetotrophic Diseases, 265.
Wilson quotients (Vandiver), 332.
WIPF, L., and SHACKELFORD, R. M. Chromosomes of a Fox Hybrid (*Alopex-Vulpes*), 468.
WOLFOWITZ, J. See Wald, A., 99.
WRIGHT, F. E. National Academy of Sciences (Distribution and Ages of Membership), 117.

X-ray induced mutation (Kerschner), 647.
X-ray induced rearrangements (Giles and Riley), 640.

YANOFSKY, C. See Bonner, D. M., 576.
Y-chromosome and mutation (Kerschner), 647.
Yeast, genetics of (Pomper and Burkholder), 456, 712.
Yeast hybrids (Lindegren and Lindegren), 23.

ZARISKI, O. A Simple Analytical Proof of a Fundamental Property of Birational Transformations, 62.
*Zea Mays* (Everett), 628.
Zermelo's axioms for sets (Wang), 150.
Zeros of derivatives (Rådström), 399.
Zeta function (Chowla and Selberg), 371
Zygospore germination (Starr), 453.